JMIRx Med

Overlay journal for preprints with post-review manuscript marketplace
Volume 5 (2024) ISSN 2563-6316 Editor in Chief: Edward Meinert, MA (Oxon), MSc, MBA, MPA, PhD, CEng, FBCS, EUR ING

Contents

Corrigenda and Addenda

Peer-Review Reports

Authors’ Response To Peer Reviewss

Original Papers

Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis

Aleksey Reshetnikov[1,*], PhD; Natalia Shaikhattarova[1,*]; Margarita Mazurok[1,*]; Nadezhda Kasatkina[2]

[1]

[2]

[*]these authors contributed equally

Corresponding Author:
Margarita Mazurok

Related Articles:

Companion article: https://preprints.jmir.org/preprint/56759

Companion article: https://www.medrxiv.org/content/10.1101/2024.01.11.24301001v1

Companion article: https://med.jmirx.org/2024/1/e60329

Companion article: https://med.jmirx.org/2024/1/e62676

Abstract

Background: Information about the range of Hounsfield values for healthy teeth tissues could become an additional tool in assessing dental health and could be used, among other data, for subsequent machine learning.

Objective: The purpose of our study was to determine dental tissue densities in Hounsfield units (HU).

Methods: The total sample included 36 healthy children (n=21, 58% girls and n=15, 42% boys) aged 10-11 years at the time of the study. The densities of 320 teeth tissues were analyzed. Data were expressed as means and SDs. The significance was determined using the Student (1-tailed) *t* test. The statistical significance was set at $P<.05$.

Results: The densities of 320 teeth tissues were analyzed: 72 (22.5%) first permanent molars, 72 (22.5%) permanent central incisors, 27 (8.4%) second primary molars, 40 (12.5%) tooth germs of second premolars, 37 (11.6%) second premolars, 9 (2.8%) second permanent molars, and 63 (19.7%) tooth germs of second permanent molars. The analysis of the data showed that tissues of healthy teeth in children have different density ranges: enamel, from mean 2954.69 (SD 223.77) HU to mean 2071.00 (SD 222.86) HU; dentin, from mean 1899.23 (SD 145.94) HU to mean 1323.10 (SD 201.67) HU; and pulp, from mean 420.29 (SD 196.47) HU to mean 183.63 (SD 97.59) HU. The tissues (enamel and dentin) of permanent central incisors in the mandible and maxilla had the highest mean densities. No gender differences concerning the density of dental tissues were reliably identified.

Conclusions: The evaluation of Hounsfield values for dental tissues can be used as an objective method for assessing their densities. If the determined densities of the enamel, dentin, and pulp of the tooth do not correspond to the range of values for healthy tooth tissues, then it may indicate a pathology.

(JMIRx Med 2024;5:e56759) doi:10.2196/56759

KEYWORDS

density; teeth; tooth; dental; dentist; dentists; dentistry; oral; tissue; enamel; dentin; Hounsfield; pathology; pathological; radiology; radiological; image; images; imaging; teeth density; Hounsfield unit; diagnostic imaging

Introduction

Healthy hard and soft dental tissues determine the quality of human life. Nowadays, there are various methods of clinical, laboratory, and instrumental studies that allow us not only to assess the initial condition of hard and soft tooth tissues but also to evaluate their change during therapeutic and preventive procedures [1,2]. Dynamic monitoring of dental tissue condition is required in trauma, after transplantation, and during therapeutic and preventive procedures [3-5]. It is especially important in children with metabolic diseases, genetic abnormalities, and special needs [6-8]. The emergence of innovative diagnostic methods provides dentists with new opportunities to assess dental health, especially in the early

stages of pathological changes that are not visible to the eye. Recently, cone-beam computed tomography (CBCT) has been widely used in dentistry [9]. Unlike traditional orthopantomograms, CBCT allows the clinician to analyze tissue density using Hounsfield units (HU) [10,11]. Information about the range of Hounsfield values for healthy teeth tissues could become an additional tool in assessing dental health, age estimation [12], and anatomy [13] and could be used, among other data, for subsequent machine learning [14]. The results of earlier studies do not provide convincing data on the range of Hounsfield values for healthy dental tissues in children of a certain age group [15,16]. Our study is aimed at establishing the Hounsfield values of dental tissue density in children in the same age group.

Methods

Ethical Considerations

The study was conducted at Resto Dental Clinic Ltd, Izhevsk, Russia, from January 2021 to January 2023. The study protocol complied with the principles outlined in the Declaration of Helsinki of the World Health Organization and was approved by the Ethics Committees at Resto Dental Clinic Ltd (protocol 07; December 22, 2020). Informed consent was obtained from the parents or legal guardians of all children in the study.

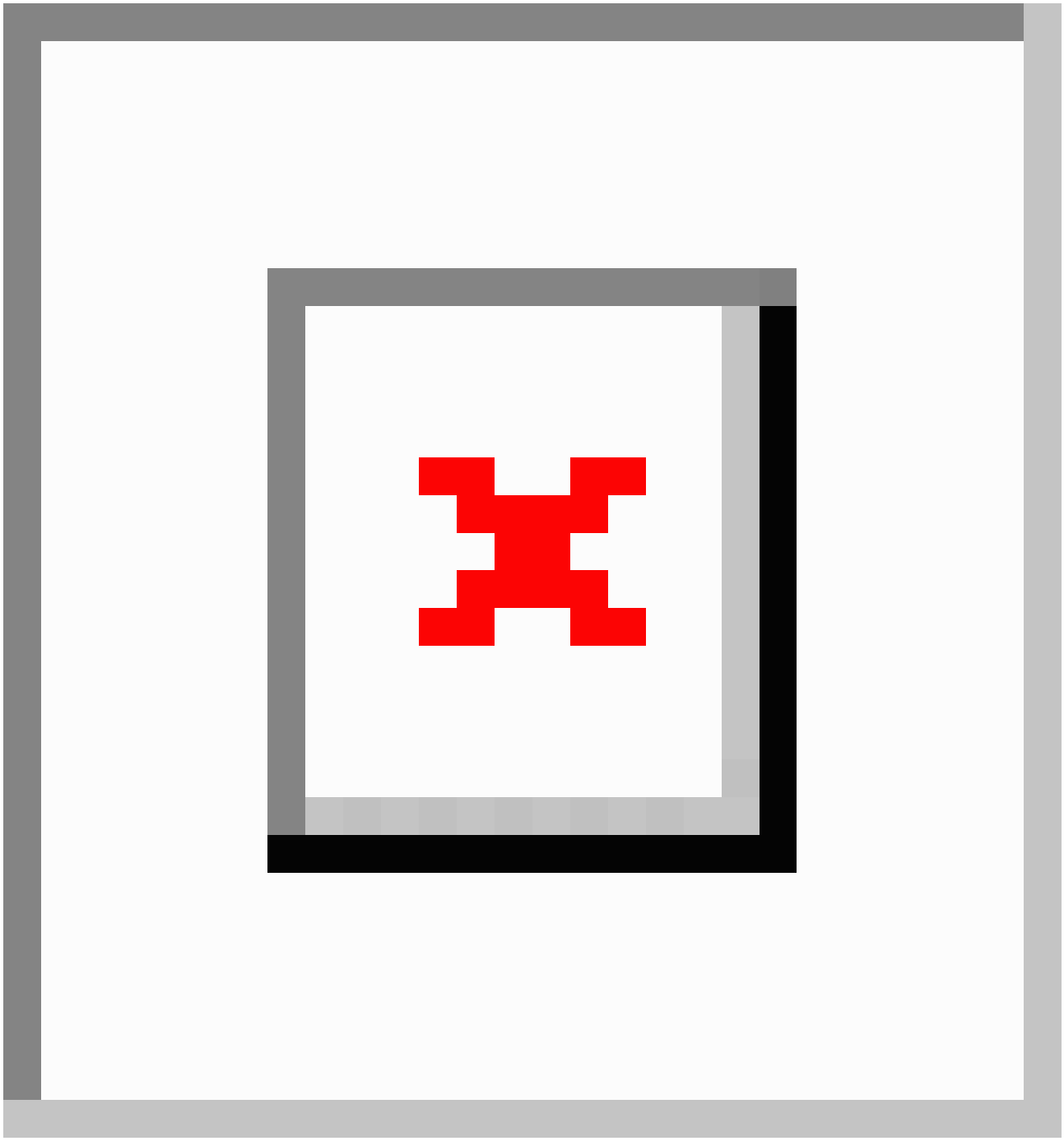
Participants

The study included 36 children aged 10 - 11 years of both genders. The criteria for including children in the study were (1) the presence of medical indications for CBCT (malocclusion and dental structural anomalies in the primary and permanent dentition, dental trauma, or anomalies in dental position), (2) aged over 10 years, (3) consent to the study, and (4) the absence of genetic anomalies and concomitant diseases.

Procedures

This study was not a randomized controlled trial and was therefore not registered at ClinicalTrials.gov. Before CBCT, all participants underwent a clinical study with a visual-tactile method. CBCT studies were performed using a PlanmecaProMax 3D tomograph (Planmeca Oy) with scanning parameters of 88 kV, 5 mA, and 15 seconds. Only 1 expert clinician performed the measurements. PlanmecaRomexis 5.2.R 24.10.18 software (Planmeca Oy) was used to analyze the data obtained. The average dental tissue density was determined over an area of 1 mm^2. Teeth tissues of the upper and lower jaws that were selected for the study included first permanent molars, permanent central incisors, second primary molars, tooth germs of second premolars, second premolars, second permanent molars, and tooth germs of second permanent molars. Teeth enamel and dentin densities were measured in HU on the incisor or occlusal surface (enamel 1 and dentin 1) and proximal surface (enamel 2 and dentin 2). Pulp density was measured in its central area (Figure 1).

Figure 1. Measuring the radiodensity (in Hounsfield units) of a first permanent molar of a boy (10 years old) under cone-beam computed tomography. H: height; W: width.



Data Analysis

Data were expressed as means and SDs. The significance was determined using the Student (1-tailed) *t* test. The statistical significance was set at $P<.05$.

Results

Baseline Characteristics

The total sample consisted of 36 healthy children (n=21, 58% girls and n=15, 42% boys) aged 10 - 11 years at the time of the study. The densities of 320 teeth tissues were analyzed: 72 (22.5%) first permanent molars, 72 (22.5%) permanent central incisors, 27 (8.4%) second primary molars, 40 (12.5%) tooth germs of second premolars, 37 (11.6%) second premolars, 9 (2.8%) second permanent molars, and 63 (19.7%) tooth germs of second permanent molars.

Dental Tissue Densities

The analysis of the data showed that tissues of healthy teeth in children have different density ranges: enamel, from mean 2954.69 (SD 223.77) HU to mean 2071.00 (SD 222.86) HU; dentin, from mean 1899.23 (SD 145.94) HU to mean 1323.10 (SD 201.67) HU; and pulp, from mean 420.29 (SD 196.47) HU to mean 183.63 (SD 97.59) HU. The statistical analysis did not reveal any significant relationships between Hounsfield values and demographic data (gender). Therefore, the densities of the

tissues of the maxilla and mandible teeth were compared. Detailed data are presented in Table 1.

The tissues (enamel and dentin) of permanent central incisors in the mandible and maxilla had the highest mean densities. The enamel and dentin densities of the second primary molars were significantly lower than those for second permanent molars and tooth germs of second permanent molars (all $P<.05$).

Table . Dental tissue densities of healthy children in Hounsfield units.

Teeth tissues and jaw		Enamel 1	Enamel 2	Dentin 1	Dentin 2
First permanent molars					
	Maxilla (n=36), mean (SD)	2426.28 (168.41)	2358.81 (219.60)	1561.17 (143.59)	1584.11 (137.17)
	Mandible (n=36), mean (SD)	2414.53 (194.85)	2336.39 (171.98)	1537.50 (150.25)	1487.19 (189.15)
	t test (*df*)[a]	0.2699 (70)	0.4756 (70)	0.6738 (70)	2.4543 (70)
	t critical value[b]	2.0003	2.0003	2.0003	2.0003
Permanent central incisors					
	Maxilla (n=36), mean (SD)	2954.69 (223.77)	2592.54 (186.54)	1796.40 (163.39)	1791.91 (127.94)
	Mandible (n=36), mean (SD)	2984.20 (223.44)	2552.37 (186.85)	1899.23 (145.94)	1871.69 (98.81)
	t test (*df*)	0.5521 (70)	0.9001 (70)	2.7769 (70)	2.9202 (70)
	t critical value	2.0003	2.0003	2.0003	2.0003
Second primary molars					
	Maxilla (n=16), mean (SD)	2141.75 (246.70)	2228.53 (160.24)	1428.06 (203.41)	1413.21 (145.79)
	Mandible (n=11), mean (SD)	2227.09 (115.66)	2071.00 (222.86)	1323.10 (201.67)	1434.50 (144.22)
	t test (*df*)	1.2045 (25)	1.9976 (25)	1.2867 (25)	0.355 (25)
	t critical value	2.0595	2.0639	2.0639	2.0739
Tooth germs of second premolars					
	Maxilla (n=21), mean (SD)	2449.71 (181.11)	2509.62 (221.56)	1576.48 (126.62)	1649.71 (128.85)
	Mandible (n=19), mean (SD)	2583.68 (134.75)	2611.32 (181.89)	1666.42 (138.10)	1695.74 (108.76)
	t test (*df*)	2.6698 (38)	1.5923 (38)	2.1394 (38)	1.2245 (38)
	t critical value	2.0211	2.0211	2.0211	2.0211
Second premolars					
	Maxilla (n=20), mean (SD)	2220.58 (190.65)	2301.32 (193.38)	1417.30 (119.57)	1507.85 (171.50)
	Mandible (n=17), mean (SD)	2336.94 (218.79)	2348.18 (103.87)	1337.00 (170.81)	1375.18 (126.17)
	t test (*df*)	1.7094 (35)	0.9365 (35)	1.6285 (35)	2.7042 (35)
	t critical value	2.0211	2.0211	2.0211	2.0211
Second permanent molars					
	Maxilla (n=4), mean (SD)	2350.00 (49.02)	2403.50 (101.93)	1569.00 (88.75)	1523.25 (91.31)
	Mandible (n=5), mean (SD)	2293.40 (131.28)	2174.40 (145.79)	1443.00 (70.81)	1327.60 (121.99)
	t test (*df*)	0.8897 (7)	2.7734 (7)	2.3111 (7)	2.7502 (7)
	t critical value	2.3646	2.3646	2.3646	2.3646
Tooth germs of second permanent molars					
	Maxilla (n=32), mean (SD)	2359.03 (169.39)	2403.16 (209.89)	1527.03 (121.39)	1519.66 (105.13)
	Mandible (n=31), mean (SD)	2356.52 (148.88)	2499.97 (178.51)	1527.81 (128.91)	1554.37 (120.88)

Teeth tissues and jaw		Enamel 1	Enamel 2	Dentin 1	Dentin 2
	t test (*df*)	0.0625 (61)	1.9741 (61)	0.0247 (61)	1.2031 (61)
	t critical value	2.0423	2.0423	2.0423	2.0423

[a]The differences are significant when *t* test value>*t* critical value at *P*=.05.

[b]*t* critical value at *P*=.05.

Discussion

Principal Findings

The densities of dental tissues are an independent sign of their health. With the development of computed tomography and software, clinicians acquired an additional tool for analyzing the density of dental tissues [11]. The determination of Hounsfield values of dental tissues using CBCT can be used as an objective method for assessing their densities in people of different age groups. We obtained the measurements of tissue densities of healthy teeth in children aged 10 - 11 years.

Previous studies of extracted teeth using microcomputed tomography showed uneven distribution of enamel and dentin densities in different areas of the tooth [14]. Yavuz et al [15] confirmed this pattern in a mixed-age population using CBCT in their study. However, the densities of enamel and dentin in their study were lower than the average values obtained during our study. One of the reasons justifying this difference may be the fact that our study included children in the same age group, which may justify further studies on the dental tissue density in a population of children and adults of certain age groups. The obtained densities for the tissues of teeth germs indicate that they correspond to the densities of permanent teeth and exceed similar indicators of the densities of primary teeth tissues.

We believe that further research on the density range for healthy and pathologically altered dental tissues, as well as study standardization, can help clinicians improve the accuracy of screening and optimize subsequent monitoring of the effectiveness of preventive and therapeutic procedures in the future. This study is an attempt to establish the range of Hounsfield values for healthy maxillary and mandibular dental tissues in children of a certain age group. The data obtained revealed the densities for enamel, dentin, and pulp for primary and permanent teeth and germs of primary teeth. Differences in the densities of specific teeth were also revealed; in particular, it was found that the enamel of the incisors had the highest density, significantly exceeding the densities of the molars. Further research on the densities of dental tissues in normal and pathological conditions seems promising, in particular for machine learning [14,17].

Limitations

A limitation of our study was that measurements were carried out by only 1 expert clinician, which eliminates an assessment of interobserver variability. The study was conducted in a population of children in the same age group. In addition, not all maxillary and mandibular teeth were included in the study. This study only obtained Hounsfield values of dental structures from 1 particularly used CBCT machine. Further studies on a larger population may be useful to improve the information content of the data.

Conclusions

The evaluation of Hounsfield values of dental tissues can be used as an objective method for assessing their densities. If the determined densities of the enamel, dentin, and pulp of the tooth do not correspond to the range of values for healthy tooth tissues, then it may indicate a pathology.

Acknowledgments

The authors acknowledge and thank Professor Aleksandr Urakov (from Izhevsk State Medical Academy) for all the support.

Conflicts of Interest

None declared.

References

1. Warreth A. Dental caries and its management. Int J Dent 2023 Jan;2023:9365845. [doi: 10.1155/2023/9365845] [Medline: 36636170]
2. Anauate-Netto C, Borelli Neto L, Amore R, Di Hipólito V, D'Alpino PHP. Caries progression in non-cavitated fissures after infiltrant application: a 3-year follow-up of a randomized controlled clinical trial. J Appl Oral Sci 2017;25(4):442-454. [doi: 10.1590/1678-7757-2016-0633] [Medline: 28877284]
3. Elgezawi M, Haridy R, Abdalla MA, Heck K, Draenert M, Kaisarly D. Current strategies to control recurrent and residual caries with resin composite restorations: operator- and material-related factors. J Clin Med 2022 Nov 7;11(21):6591. [doi: 10.3390/jcm11216591] [Medline: 36362817]
4. Abdou A, Matoug-Elwerfelli M, Nagendrababu V, Nazzal H, Duggal M. Tooth auto-transplantation: a bibliometric analysis of the top 100 most-cited articles. Dent Traumatol 2023 Feb;39(1):64-81. [doi: 10.1111/edt.12779] [Medline: 35972826]

5. Jeong JS, Pang NS, Choi Y, et al. Importance of photography education to improve image quality for accurate remote diagnoses in dental trauma patients: observational study. JMIR Mhealth Uhealth 2020 Mar 26;8(3):e15152. [doi: 10.2196/15152] [Medline: 32213475]
6. Li X, Ye X, Su J. The dental phenotype of primary dentition in SATB2-associated syndrome: a report of three cases and literature review. BMC Oral Health 2022 Nov 22;22(1):522. [doi: 10.1186/s12903-022-02594-4] [Medline: 36457071]
7. Patni T, Lee CT, Li Y, et al. Factors for poor oral health in long-term childhood cancer survivors. BMC Oral Health 2023 Feb 4;23(1):73. [doi: 10.1186/s12903-023-02762-0] [Medline: 36739372]
8. Pecci-Lloret MR, Pecci-Lloret MP, Rodríguez-Lozano FJ. Special care patients and caries prevalence in permanent dentition: a systematic review. Int J Environ Res Public Health 2022 Nov 17;19(22):15194. [doi: 10.3390/ijerph192215194] [Medline: 36429911]
9. Pauwels R, Jacobs R, Singer SR, Mupparapu M. CBCT-based bone quality assessment: are Hounsfield units applicable? Dentomaxillofac Radiol 2015;44(1):20140238. [doi: 10.1259/dmfr.20140238] [Medline: 25315442]
10. Shapurian T, Damoulis PD, Reiser GM, Griffin TJ, Rand WM. Quantitative evaluation of bone density using the Hounsfield index. Int J Oral Maxillofac Implants 2006;21(2):290-297. [Medline: 16634501]
11. Mathur VP, Dhillon JK, Logani A, Kalra G. Evaluation of indirect pulp capping using three different materials: a randomized control trial using cone-beam computed tomography. Indian J Dent Res 2016;27(6):623-629. [doi: 10.4103/0970-9290.199588] [Medline: 28169260]
12. Yousefi F, Mohammadi Y, Ahmadvand M, Razaghi P. Dental age estimation using cone-beam computed tomography: a systematic review and meta-analysis. Imaging Sci Dent 2023 Jun;53(2):91-99. [doi: 10.5624/isd.20221226] [Medline: 37405204]
13. Almeida JC, Candemil AP, Bertolini GR, et al. Cone-beam computed tomographic evaluation of the root canal anatomy of the lower premolars and molars in a Brazilian sub-population. Imaging Sci Dent 2023 Mar;53(1):77-82. [doi: 10.5624/isd.20220204] [Medline: 37006788]
14. Hung KF, Ai QYH, Leung YY, Yeung AWK. Potential and impact of artificial intelligence algorithms in dento-maxillofacial radiology. Clin Oral Investig 2022 Sep;26(9):5535-5555. [doi: 10.1007/s00784-022-04477-y] [Medline: 35438326]
15. Yavuz Y, Akleyin E, Doğan MS, Goncharuk-Khomyn M, Akkus Z. Can the density of mineralized dental tissues (dentin and enamel) be measured and compared with 3D cone beam computed tomography in cases of ectodermal dysplasia? Med Sci Monit 2022 Jul 1;28:e937003. [doi: 10.12659/MSM.937003] [Medline: 35773958]
16. Hayashi-Sakai S, Sakamoto M, Hayashi T, et al. Evaluation of permanent and primary enamel and dentin mineral density using micro-computed tomography. Oral Radiol 2019 Jan;35(1):29-34. [doi: 10.1007/s11282-018-0315-2] [Medline: 30484178]
17. He F, Ng Yin Ling C, Nusinovici S, et al. Development and external validation of machine learning models for diabetic microvascular complications: cross-sectional study with metabolites. J Med Internet Res 2024 Mar 28;26:e41065. [doi: 10.2196/41065] [Medline: 38546730]

Abbreviations

CBCT: cone-beam computed tomography
HU: Hounsfield units

Edited by E Meinert, F Wu, T Leung; submitted 25.01.24; peer-reviewed by S Gorthy, Anonymous; revised version received 11.04.24; accepted 21.04.24; published 20.06.24.

Please cite as:
Reshetnikov A, Shaikhattarova N, Mazurok M, Kasatkina N
Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis
JMIRx Med 2024;5:e56759
URL: https://xmed.jmir.org/2024/1/e56759
doi:10.2196/56759

The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence

Ramendra Pati Pandey[1,*], PhD; Himanshu[2,3,*], MSc; Gunjan[2,3], MSc; Riya Mukherjee[2,3], MSc; Chung-Ming Chang[3,4,*], PhD

1

2

3

4

*these authors contributed equally

Corresponding Author:
Chung-Ming Chang, PhD

Related Articles:

Companion article: https://preprints.jmir.org/preprint/51787

Companion article: https://med.jmirx.org/2024/1/e56047

Companion article: https://med.jmirx.org/2024/1/e56440

Companion article: https://med.jmirx.org/2024/1/e55899

Abstract

Background: Animal-assisted therapy, also known as pet therapy, is a therapeutic intervention that involves animals to enhance the well-being of individuals across various populations and settings.

Objective: This systematic study aims to assess the outcomes of animal-assisted therapy interventions and explore the associated policies.

Methods: A total of 16 papers published between 2015 and 2023 were selected for analysis. These papers were chosen based on their relevance to the research topic of animal-assisted therapy and their availability in scholarly databases. Thematic synthesis and meta-analysis were used to synthesize the qualitative and quantitative data extracted from the selected papers.

Results: The analysis included 16 studies that met the inclusion criteria and were deemed to be of moderate or higher quality. Among these studies, 4 demonstrated positive results for therapeutic mediation and one for supportive mediation in psychiatric disorders. Additionally, all studies showed positive outcomes for depression and neurological disorders. Regarding stress and anxiety, 3 studies indicated supportive mediation, while 2 studies showed activating mediation.

Conclusions: The overall assessment of animal-assisted therapy shows promise as an effective intervention in promoting well-being among diverse populations. Further research and the establishment of standardized outcome assessment measures and comprehensive policies are essential for advancing the field and maximizing the benefits of animal-assisted therapy.

(JMIRx Med 2024;5:e51787) doi:10.2196/51787

KEYWORDS

animal-assisted therapy; pet therapy; outcome assessment; policies; systematic study

Introduction

The inclusion of animals in psychological treatment is not a recent or uncommon practice. Throughout history, there has been an understanding of the positive impact animals can have on human well-being [1]. This connection between humans and animals is deeply ingrained in our collective subconscious, influencing our emotional experiences [2]. The earliest documented instance dates back to the late 18th century when animals were introduced into mental health institutions to

enhance social interaction among patients [3,4]. Today, numerous programs worldwide incorporate animals to varying degrees in their services. These programs are particularly beneficial for individuals who have experienced trauma, including those diagnosed with posttraumatic stress disorder (PTSD), schizophrenia, Alzheimer disease, autism, etc [4,5].

In the past 50 years, the field of human-animal interaction and, specifically, animal-assisted therapy (AAT) has made significant advancements and progress. AAT is a therapeutic approach that uses animals to improve overall health and well-being. It encompasses emotional, psychological, and physical interactions between individuals, animals, and the environment [6]. AAT interventions involve qualified treatment providers facilitating interactions between patients and animals with specific therapeutic goals in mind. These interventions often involve collaborative activities between human-animal teams, aiming to promote therapeutic and supportive outcomes [7]. AAT interventions contribute to individuals' well-being, supporting physical health and improving cognitive, emotional-affective, and social aspects, leading to enhanced emotional well-being, reduced anxiety, and decreased stress levels [8-10].

Research on therapies involving human-animal interaction has focused on specific animals such as dogs, cats, or horses and specific populations such as those with autism [11]. Dogs, in particular, are commonly preferred for therapy due to their exceptional bond with humans in modern times. Over thousands of years of shared evolutionary history [1], dogs have acquired adept socialization skills with humans through processes of domestication and natural selection. They have become our loyal companions, developing unique social skills for interacting with humans [12]. For instance, studies indicate that dogs possess a sensitivity to our emotional states [13] and can interpret our social cues [14], even engaging in sophisticated communication through behaviors like gaze alternation [15]. Furthermore, dogs are capable of forming intricate attachment relationships with humans, resembling the bonds found in relationships between infants and caregivers [16]. Research suggests that among the various animals involved in AAT, dogs tend to exhibit superior interactions with people compared to other species, benefiting both children and adults [6].

This systematic review and meta-analysis sheds light on the potential of animal-assisted interventions to enhance overall well-being and health. Our research aims to contribute to the growing body of evidence supporting the use of animals in therapeutic contexts and to explore the specific contexts in which these interventions are most effective. One of the unique aspects of our study is the incorporation of both quantitative and qualitative analyses to provide a comprehensive understanding of the effects of AAT. While previous research has predominantly relied on quantitative data, we believe that qualitative insights from participants who have experienced these interventions offer valuable perspectives. The special bonds formed between humans and animals are recognized as essential catalysts for transformation and are held in high regard, similar to the therapist-client relationship.

Methods

Search Strategy

The meta-analysis was carried out following the methodologies outlined in the esteemed Cochrane Handbook for Systematic Reviews of Interventions [17], and the findings were reported in compliance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [18]. To ensure comprehensive coverage, electronic databases were meticulously searched up until June 2023. A total of five English-language electronic databases, including PubMed, Web of Science, Clinical Trials, Science Direct, and Google Scholar, were meticulously explored. This thorough exploration entailed using a combination of pertinent controlled vocabulary terms (eg, Medical Subject Headings [MeSH]) and relevant free-text terms. The search strategy used can be summarized as follows: (animal assisted therapy OR animal assisted intervention OR animal assisted activity OR animal activity interaction OR animal assisted method OR animal facilitated therapy OR pet therapy OR canine assisted therapy OR dog assisted therapy) AND (quasi-experimental study OR randomized controlled trial) AND (pain OR anxiety OR depression OR blood pressure OR BP OR heart rate OR HR) AND (work-related stress OR workplace health OR employee well-being OR burnout) AND (tumor OR malignant OR carcinoma OR oncology OR hospitalization OR hospitalized patients OR inpatients). By using this extensive and refined approach, the meta-analysis aimed to capture a comprehensive body of evidence on the effects of AAI on various health outcomes.

Inclusion and Exclusion Criteria

The inclusion criteria were set based on the PICOS (Patient, Problem, or Population; Intervention; Comparison; Outcomes; and Study Design) framework: studies evaluating the effects of AAT, animal-assisted intervention, or animal-assisted activity; studies evaluating the effects of animal interactions on health and well-being (including depression, agitation, loneliness, stress, and quality of life), social interaction, engagement, physical function, behavioral symptoms, medication use, and adverse events; articles published in English; studies available in full-text format; and studies using quasi-experimental designs or randomized controlled trials. To maintain the rigor and relevance of the study, publications that lacked sufficient information regarding the therapy or did not involve an animal intervention were excluded from consideration.

Data Synthesis

In our study, we used thematic synthesis as a method to assess the eligibility and quality of the articles [19]. Each article was independently reviewed to determine its suitability for inclusion. We followed a traditional methodology for evaluating the papers that involved examining factors such as the presence of adequate control groups, control of confounders, randomization, well-described experimental design, and relevant outcome variables. Articles that met these criteria were selected and organized into a single sheet using Microsoft Office Excel (2019; Microsoft Corporation). For the included studies, we extracted and compiled various data points into a structured table. This information encompassed the author's name, country

of publication, year of publication, patient characteristics (including sample size, age, gender, and target group), type of study, study design, description of AAT, type of intervention, control group details, study duration, outcomes measured, and the authors' conclusions. To effectively manage the papers, we used Mendeley software (version 1.19.8; Elsevier).

Classification

To determine the specific contexts in which AATs are effective, we classified the interventions into three categories. First, "supportive mediation" involves AATs providing emotional and psychological support to individuals. Second, "therapeutic mediation" entails AATs addressing specific therapeutic goals and needs in a structured manner. Finally, "activating mediation" comprises AATs designed to stimulate engagement and participation in various activities or tasks.

Risk of Bias

Assessing publication bias is a crucial component in safeguarding the strength and credibility of our meta-analysis, which examines the effects of AAT on improving the well-being of individuals in diverse populations and settings. To gauge the possible influence of publication bias on our results, we applied several established techniques recommended in the field. One of these methods involved visually examining a bias risk graph for signs of asymmetry, which can be an indication of publication bias. By using these comprehensive approaches, our objective was to address any potential bias and guarantee that our meta-analysis offers an impartial synthesis of the existing evidence regarding the beneficial effects of AAT.

Results

Study Selection

The outcome of the search is depicted in Figure 1. The search process resulted in 968 unique articles after initial searches from various electronic databases like PubMed, Web of Science, Science Direct, and Google Scholar, which yielded 942 articles. An additional 26 articles were extracted from other sources. After eliminating duplicate articles, the total number of articles was reduced to 507. Subsequently, the articles were assessed based on their title and abstract to determine eligibility. Among the initial pool, 389 articles were excluded as they did not meet the eligibility criteria, mainly due to the lack of relevance to AAT. After reading the full text of the remaining articles, 102 more articles were excluded. Of these 102 articles, 60 did not meet the inclusion criteria, and 42 were excluded due to being classified as nonoccupational mixed groups or having unrepresentative results. Finally, a total of 16 studies that met the inclusion criteria were included in the final analysis. The findings and details of these 16 studies are summarized in Tables 1 and 2.

Figure 1. Literature screening flowchart (PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses). AAT: animal-assisted therapy.

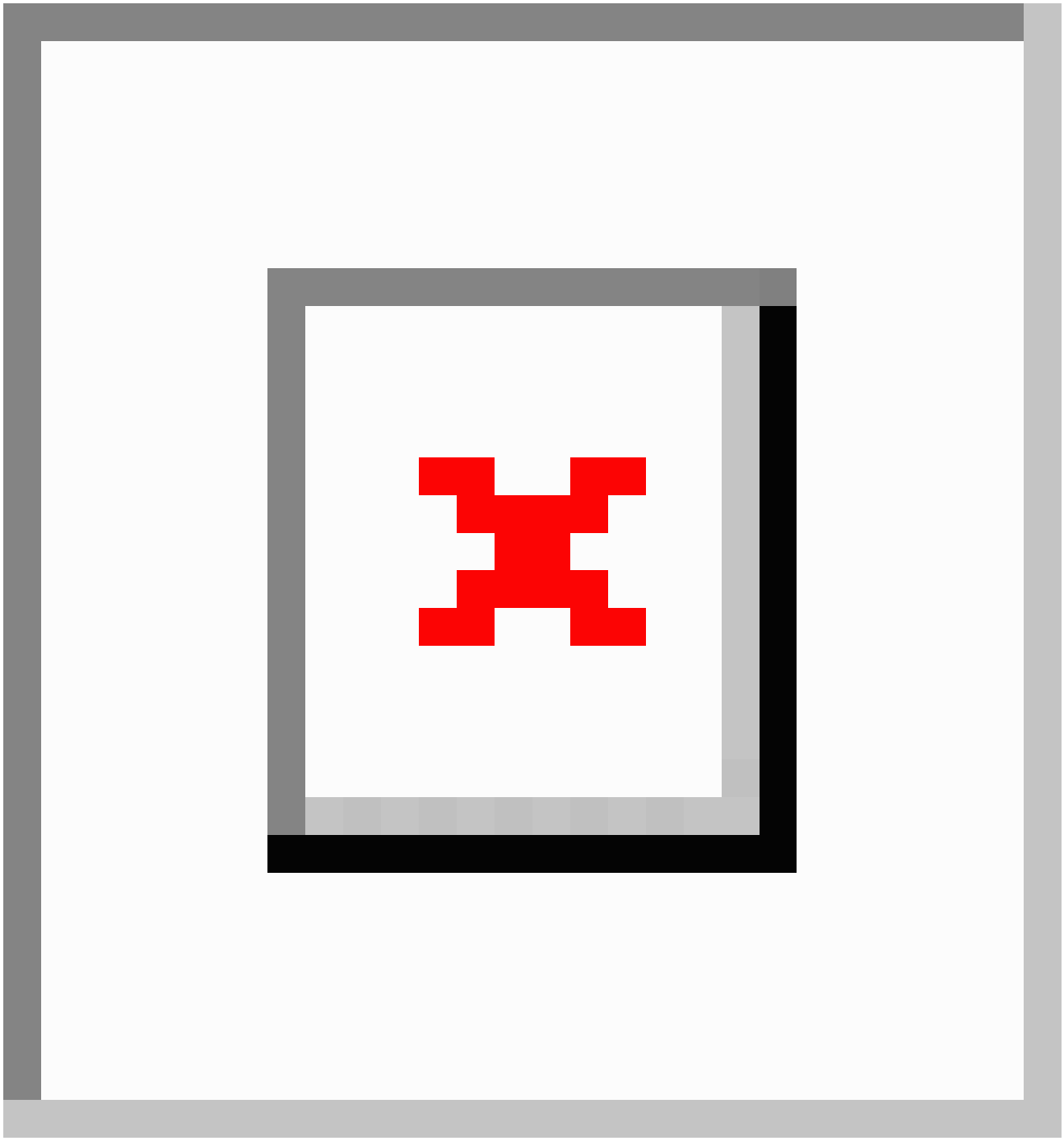


Table . Summary of outcomes from studies included.

Author; year	Type of disorder	Mediation	Study outcome of AAT[a]	Conclusion
Shih and Yang [20]; 2023	Psychiatric disorder	Therapeutic	Mental Health–Social Functioning Scale, Social Adaptive Function Scale, Taiwanese version of the World Health Organization Quality of Life	Social functioning was significantly higher in the experimental group; quality of life improved
Chen et al [21]; 2022	Psychiatric disorder	Therapeutic	Montreal Cognitive Assessment, chair stand test, Timed Up and Go, 5-min walk test, Assessment of Communication and Interaction Skills	Significant improvements in communication and interpersonal skills; improved lower extremity strength and social functions
Allen et al [22]; 2022	Stress and anxiety	Activating	Service Satisfaction Scale, Posttraumatic Stress Disorder Reaction Index for DSM-5, Strengths and Difficulties Questionnaire, Moods and Feelings Questionnaire, Screen for Child Anxiety Related Disorders	Improvements in caregiver-reported PTSD[b] symptoms, internalizing concerns, and externalizing problems
Chen et al [23]; 2021	Psychiatric disorder	Therapeutic	Positive and Negative Syndrome Scale, Depression Anxiety Stress Scales Assessment, and Chinese Happiness Inventory	Decrease in stress in the AAT group more than in the control group
Anderson and Brown [24]; 2021	Stress and anxiety	Supportive	STAI[c]	Decreased anxiety in a convenience sample
Thakkar et al [25]; 2020	Stress and anxiety	Activating	Modified faces version of the Modified Child Dental Anxiety Scale	Effective behavior management strategy for children
Santaniello et al [26]; 2020	Neurological disorder	Therapeutic	MMSE[d], GDS[e]	AAT showed an improvement in both cognitive function and mood
Brown et al [27]; 2019	Psychiatric disorder	Therapeutic	Wilcoxon signed rank test	Positive therapeutic impact on patients and staff in acute care psychiatric units, promoting positive mood and emotions
Hinic et al [28]; 2019	Stress and anxiety	Activating	STAI for Children	Reduce anxiety in hospitalized children and enhanced family satisfaction
Ginex et al [29]; 2018	Depression	Supportive	Patient Health Questionnaire–4	Promotes a healing environment for patients that involves a holistic and humanistic perspective
Priyanka MB [30]; 2018	Neurological disorder	Therapeutic	Autism Treatment Evaluation Checklist and semistructured interview for 15 min	Improved expression and communication skills when interacting with the dog, as well as noticeable enhancements in social and motor abilities
McCullough et al [31]; 2017	Stress and anxiety	Supportive	STAI	Help in reducing stress and anxiety levels
Branson at al [32]; 2017	Stress and anxiety	Supportive	STAI for Children, Positive and Negative Affect Schedule for Children	Increase positive feelings in hospitalized children

Author; year	Type of disorder	Mediation	Study outcome of AAT[a]	Conclusion
Nurenberg et al [5]; 2015	Psychiatric disorder	Therapeutic	Equine Assisted Growth and Learning Association, equine-assisted psychotherapy, canine-assisted psychotherapy	Effective therapeutic modality for long-term psychiatric patients at risk of violence
Stefanini et al [33]; 2015	Psychiatric disorder	Supportive	Children Global Assessment Scale	Significant improvements in overall functioning, a decrease in the need for specialized care, and an increase in regular school attendance
Menna et al [34]; 2015	Neurological disorder	Therapeutic	MMSE, 15-item GDS	Improved cognition and mood through repeated multimodal stimulation

[a]AAT: animal-assisted therapy.

[b]PTSD: posttraumatic stress disorder.

[c]STAI: State-Trait Anxiety Inventory.

[d]MMSE: Mini-Mental State Examination.

[e]GDS: Geriatric Depression Scale.

Table . Study characteristics.

Authors; year	Country	Study design	Patients					Study conditions	AAT[a] description	Intervention	Control
			Number	Gender, n		Age (years), mean	Target group				
				Male	Female						
Shih and Yang [20]; 2023	Taiwan	Randomized controlled	90 (AAT group: 45; control group: 45)	45	45	50.2	Patients with schizophrenia	1-h therapy session once a week for 12 wk	Physical contact, brushing, playing, walking, and sitting	Dog-assisted therapy	Regular therapy
Chen et al [21]; 2022	Taiwan	Randomized controlled	40 (AAT group: 20; control group: 20)	18	22	54.6	Patients with schizophrenia	1-h therapy session once a week for 12 wk	15-min warm-up session like greetings, introduction, and season orientation; 45-min therapeutic activity, physical activities, or cognitive activities	Dog-assisted therapy	Regular therapy
Allen et al [22]; 2022	United States	Randomized controlled	33 (AAT group: 17; control: 16)	11	22	15	Abused youth with PTSD[b]	12 sessions, each lasting 90 min	Physical contact, petting	Dog present at the time of questionnaire	Dog was not present
Chen et al [23]; 2021	Taiwan	Randomized controlled	40 (AAT group: 20; control group: 20)	18	22	55.3	Patients with schizophrenia	1-h therapy session once a week for 12 wk	Petting, massaging, and playing with the dog (ball, loop, game)	Dog-assisted therapy	Regular therapy
Anderson and Brown [24]; 2021	United States	Randomized controlled	89 (AAT group: 45; control group: 44)	9	80	22.6	Nursing students anxiety	35-45 min before exam	Unstructured	Dog-assisted therapy	No interaction with dog
Thakkar et al [25]; 2020	India	Randomized controlled	100 (AAT group: 50; control group: 50)	44	56	7.5	Children who underwent dental assessment	Play with dog for 10-15 min in operatory	Petting and conversation with dog handler	Dog present during dental procedure	Dog was not present
Santaniello et al [26]; 2020	Italy	Randomized controlled	96 (AAT group- 65; control group-31)	23	75	75.8	Patients with Alzheimer disease	45-min therapy session once a week for 6 mo	Physical contact, brushing, playing, walking, and sitting	Dog present during therapy	Dog was not present during therapy

Authors; year	Country	Study design	Patients					Study conditions	AAT[a] description	Intervention	Control
			Number	Gender, n		Age (years), mean	Target group				
				Male	Female						
Brown et al [27]; 2019	United States	Time series and daily announcement	152 (adult inpatient unit: 84; adolescent inpatient unit: 68)	28; 18	56; 50	58	—[c]	Weekly dog visit	Interaction with therapy dog and its handler	Dog-assisted therapy	Regular therapy
Hinic et al [28]; 2019	United States	Purposive sampling	93 (AAT group: 50; control group: 43)	40	53	10.5	Hospitalized children	10-min pet therapy per visit	Dog present at the time of questionnaire; petting the dog	Dog-assisted therapy	No interaction with dog
Ginex et al [29]; 2018	United States	Randomized controlled	100 (AAT group: 50; control group: 50)	48	52	58	Patients of oncological surgical unit	4 d a week for 6 wk	Unstructured	Dog-assisted therapy	Regular therapy
Priyanka MB [30]; 2018	India	Purposive sampling	6	—	—	8.6	Children with autism	10 d for 12 wk	Petting and brushing the dog	Dog-assisted therapy	Regular therapy
McCullough et al [31]; 2017	United States	Randomized controlled	100 (AAT group: 60; control group: 46)	57	49	4.5	Patients with cancer	10-20 min per session	Petting and brushing the dog	Dog-assisted therapy	Regular therapy
Branson et al [32]; 2017	United States	Randomized controlled	48 (AAT group: 24; control group: 24)	24	24	13.4	Hospitalized children	10-min interaction in waiting room	Unstructured	Dog interaction	No interaction
Nurenberg et al [5]; 2015	United States	Randomized controlled	90	57	33	44.7	Chronic psychiatric inpatients	Weekly group session	Unstructured	Dog interaction	No interaction
Stefanini et al [33]; 2015	Italy	Randomized controlled	34 (AAT group: 17; control group: 17)	18	16	15	Acute mental patients	45-min therapy session once a week for 3 mo	Play activities, physical contact, grooming, cleaning	Dog interaction	No interaction
Menna et al [34]; 2015	Italy	Divided according to conditions	50 (AAT group: 40; control group: 10)	13	37	75.1	Patients with Alzheimer disease	45-min therapy session once a week for 6 mo	15-min reintroduction to dog; 20-min structured activity; 10-min same ending activity	Dog interaction	No interaction

[a]AAT: animal-assisted therapy.

[b]PTSD: posttraumatic stress disorder.

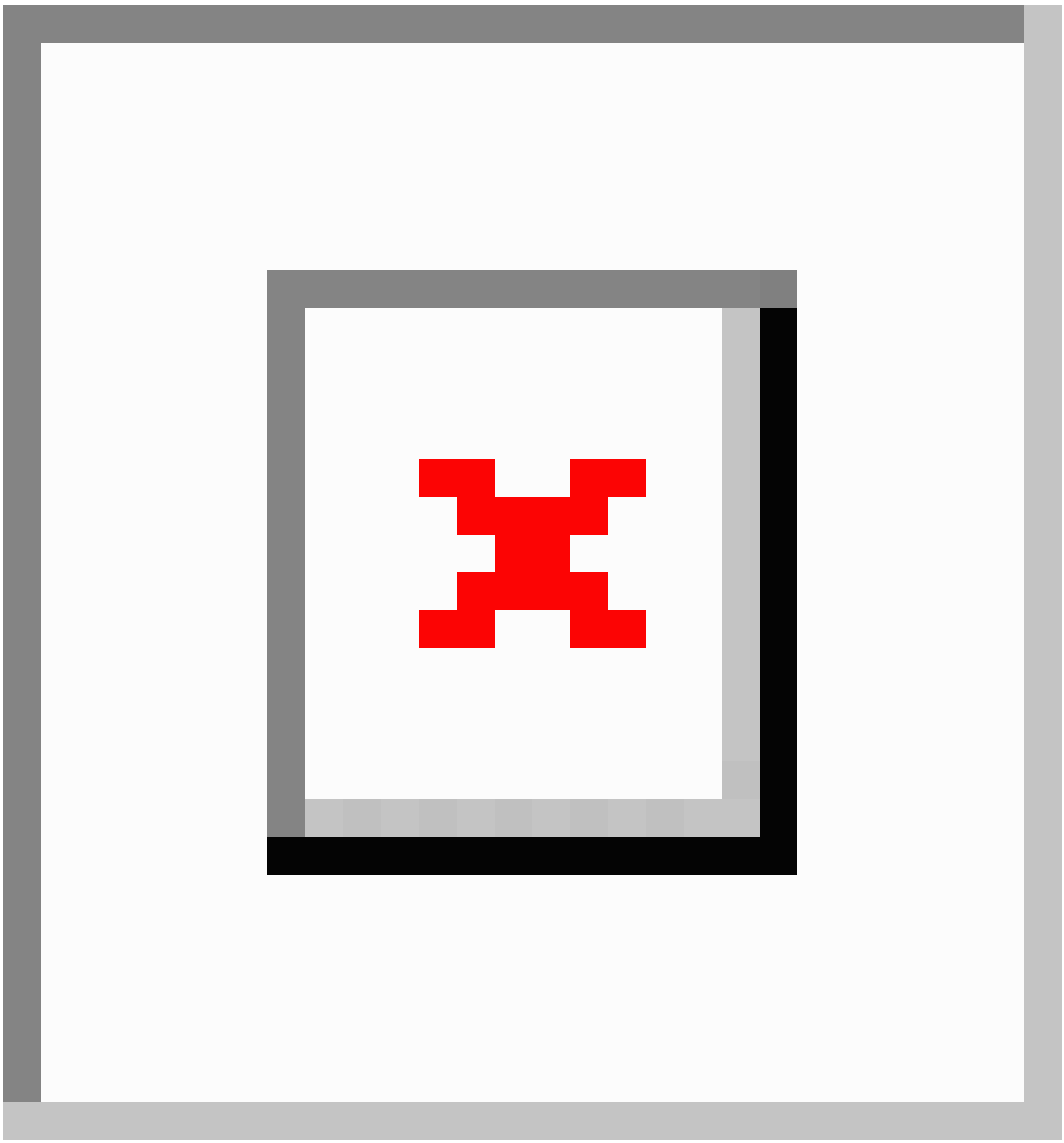
[c]Not available.

Risk of Bias Determination

Figure 2 [5,20-34] provides a comprehensive evaluation of the overall risk and individual biases in each of the studies included. All the researchers carried out assessments to determine the likelihood of bias, and the results of these assessments showed a remarkable level of consistency across all the investigations.

The data depicted in Figure 2 indicate that the articles authored by Hinic et al [28], Priyanka MB [30], and Menna et al [34] exhibited a high risk of bias. Additionally, the study conducted by Anderson and Brown [24] indicated a potential risk of bias. This consistent and rigorous approach enhances the confidence in the research paper's results, underscoring the reliability of the reported biases and their impact on the study's outcomes.

Figure 2. Risk of bias for the selected studies.



General Characteristics

To provide a concise summary of the selected studies, we have compiled an overview in Table 2. It presents key information from each study, allowing for a quick and comprehensive understanding of the research landscape.

Quantitative Study

The studies analyzed in this research were conducted between 2015 and 2023, resulting in a total of 16 [5,20-34] included papers. These studies were carried out in various countries, including the United States (n=8), Taiwan (n=3), India (n=2), and Italy (n=3). Among the included studies, approximately 12

of 16 (75%) used a randomized controlled trial design. Two studies used conditional controlled designs, while 1 study followed a time series and daily announcement approach. The number of patients enrolled in the studies varied, ranging from 6 to 152 individuals. Specifically, 5 studies involved 0-40 patients, 2 studies included 41-80 patients, 8 studies comprised 81-120 patients, and 1 study encompassed 121-160 patients. The selected studies covered a diverse range of populations, with 7 studies focusing on children and adolescents from various disciplines including child psychology, psychiatry, and pediatrics. The same number of studies involved adult patients covering a range of fields such as general medicine, mental health, and geriatrics, and 2 studies specifically targeted older people, contributing to the fields of gerontology and geriatric medicine. In terms of gender distributions, women were more prominently represented, with ≥50% female participants in 13 of the 16 studies (Table 2).

The interventions included in the studies were described using various terms such as pet encounter therapy, pet-facilitated therapy, pet-assisted living, animal-assisted intervention, AAT, animal-assisted activity, or simply dog visits/therapy. All of the studies incorporated dogs as the primary intervention. In terms of the duration of the interventions, the 7 studies had varying time periods per visit [5,24,25,27,28,31,32]. Five studies had interventions lasting for 12 weeks [20,21,23,30,33] and one for 6 weeks [29], while 2 studies had longer intervention periods of 6 months [26,34]. One study did not explicitly mention the duration of the intervention [22]. The majority of the studies used a one-to-one approach in delivering the intervention, emphasizing individual interactions between participants and the dogs. However, 1 study was conducted in a group setting [5]. Most studies actively encouraged touching and interaction with the animals, while in 2 studies, the interaction was described as unstructured [24,32].

Qualitative Study

The description of selected studies for qualitative analysis is presented in Tables 1 and 2. Studies reported on patients' experiences with animal-assisted interventions such as dog therapy or animal visits. Thematic analysis was used to identify recurring themes and extract meaningful insights from the type of disorder. Participants described the animals as a source of comfort, providing emotional support and reducing stress and anxiety. The interactions with the animals were reported to have a soothing effect and helped individuals cope with their challenges and emotional difficulties.

The qualitative analysis shed light on the subjective experiences and perceptions of individuals participating in the interventions. It provided valuable insights into the emotional, social, and therapeutic benefits associated with animal-assisted interventions, highlighting the potential of these interventions to enhance well-being and quality of life. The selected studies on psychiatric disorders predominantly focused on schizophrenia, with 5 studies specifically addressing this condition in adults. Additionally, 1 study explored acute mental disorders in children [33]. Six studies were dedicated to investigating stress and anxiety, targeting various populations such as children undergoing physical examinations, children with PTSD, patients with cancer, and nursing students. Three studies examined neurological disorders, including 1 study involving children with autism [30] and 2 studies involving older individuals with Alzheimer disease [26,34]. In 1 study, the intervention aimed to reduce depression among patients undergoing oncology surgery [25].

Outcomes

The number of studies with at least one statistically significant positive outcome measure, divided by patient condition and intervention category, is presented in Table 3. The study aimed to comprehensively evaluate mental health, social functioning, and overall quality of life, taking into account various parameters specific to each measurement scale, for example, generic health-related quality of life measures like Posttraumatic Stress Disorder Reaction Index for DSM-5, State-Trait Anxiety Inventory (STAI) for Children, Patient Health Questionnaire–4 (PHQ-4), Mini-Mental State Examination (MMSE), and 15-item Geriatric Depression Scale (GDS), and general functional measures such as Mental Health–Social Functioning Scale, Social Adaptive Function Scale, chair stand test, Timed Up and Go, Assessment of Communication and Interaction Skills, etc.

Table . Number of studies classified based on the condition, type of intervention, and the presence of positive outcome.

Condition	Supportive mediation, n		Therapeutic mediation, n		Activating mediation, n	
	Yes	No	Yes	No	Yes	No
Psychiatric disorder	1	0	4	1	0	0
Neurological disorder	0	0	3	0	0	0
Stress and anxiety	3	0	0	0	2	1
Depression	1	0	0	0	0	0

Psychiatric Disorder

All 6 trials that focused on psychiatric disorders were categorized as AAT and involved interventions with dog therapy. Among these studies, 5 were conducted using a randomized controlled design, while 1 study used a time series design with randomized daily announcements within a pre-post experimental framework. One study specifically examined patients in child and adolescent psychiatry [33], while the remaining 5 studies focused on adult psychiatry patients [5,20,21,23,27]. The duration of the AAT programs varied, with some studies consisting of 12-week programs in different

settings, while 2 studies provided weekly therapy sessions without specifying the intervention period [5,27].

Each of the 6 conducted studies involved a comparison between an intervention group receiving a specific therapy and a control group that did not participate in any related activities. Notably, the 5 studies specifically targeted middle-aged and older patients diagnosed with chronic schizophrenia. The results of these studies consistently demonstrated significant improvements in various areas, including reductions in psychiatric symptoms, enhanced social functioning, improved quality of life, enhanced cognitive function, increased agility and mobility, and decreased stress levels. These outcomes were measured using a variety of scales and assessment tools [5,20,21,23,33]. In a study conducted by Brown et al [27], the focus was on examining the impact of mood states and feelings among patients and staff in inpatient psychiatric units. The researchers observed significant changes in mood before and after sessions involving therapy dogs. Specifically, negative moods decreased, while positive moods, such as feelings of happiness, relaxation, and calmness, increased. These changes were measured using the visual analog mood scale [27]. Overall, these findings highlight the efficacy of AAT in positively impacting the well-being and overall functioning of individuals with psychiatric disorders.

Neurological Disorder

Among the studies that focused on neurological disorders, 3 used dog therapy as an intervention. One of these studies used a randomized controlled design [26], while the other 2 studies used purposive sampling based on the patients' conditions. One study specifically targeted children and adolescents with autism, while the other 2 studies focused on older patients with Alzheimer disease. The duration of the AAT programs varied, ranging from 3 to 6 months.

In each of the 3 conducted studies, the intervention group was compared to a control group that did not participate in any activities to assess the outcomes of the therapy. Priyanka MB's [30] study focused on children with autism and observed that engaging with a therapy dog, such as brushing the dog and attempting to draw and write for the dog, led to enhanced social and motor skills. Additionally, the children experienced a sense of relaxation and calmness in the presence of the dog. The studies conducted by Menna et al [34] and Santaniello et al [26], focusing on older patients with Alzheimer disease over 6 months, have shown promising results. Menna et al's [34] study demonstrated the applicability and effectiveness of AAT interventions in stimulating cognition and improving mood. The interventions involved repeated multimodal stimulation, including verbal, visual, and tactile approaches. Similarly, Santaniello et al's [26] study also revealed improvements in both cognitive function and mood in the AAT group, as measured by changes in the MMSE and GDS. Overall, these studies indicate that nonpharmacological therapies, particularly AAT, have the potential to reduce symptoms associated with neurological disorders.

Stress and Anxiety

The 6 trials that specifically addressed stress and anxiety used AAT interventions involving dog therapy. These studies exclusively targeted children and adolescents, using a randomized controlled design. In 5 of the studies, the therapy sessions lasted between 10-45 minutes, while 1 study did not specify the duration of the intervention period.

The study conducted by Allen et al [22] focused on youths who had experienced abuse and were diagnosed with PTSD. The results revealed that the group receiving the intervention showed greater improvements in caregiver-reported symptoms of PTSD, internalizing concerns, and externalizing problems compared to the control group [22]. In a study by Anderson and Brown [24] involving nursing students, the intervention group experienced interactions with dogs before testing. This interaction served as a stress reliever for the students, resulting in a decrease in anxiety as measured by the STAI. Thakkar et al [25] conducted a study on children who were undergoing dental assessments. The findings indicated that the intervention group showed a significantly greater anxiety reduction compared to the control group, as measured by the modified faces version of the Modified Child Dental Anxiety Scale. In the studies conducted by Hinic et al [28] and Branson et al [32], dog therapy was provided to children who were hospitalized, and their anxiety levels were assessed before and after the intervention. The results from the STAI for Children suggested that brief pet therapy visits served as a tool to decrease anxiety in children who were hospitalized and promote family satisfaction. McCullough et al [31] conducted a study where the intervention group participated in dog therapy, while the control group received standard care at the hospital. The findings demonstrated the applicability and effectiveness of AAT interventions in reducing stress and anxiety levels in patients with cancer.

Overall, when considering the results of all these studies, it becomes evident that each one exhibited at least one statistically significant positive effect. When these findings are examined collectively, they provide compelling evidence to suggest that particular modalities of AAT hold substantial promise in terms of reducing stress levels and fostering a positive impact on individuals' overall mood and well-being.

Depression

Ginex et al [29] conducted a study to explore the impact of a dog-assisted intervention on an inpatient surgical oncology unit. The study used a randomized controlled design, with patients in the intervention group receiving therapy 4 days per week throughout the study period. In contrast, the control group underwent physical therapy without any modifications to their normal routine. Patients in the intervention group reported a significant decrease in depression and anxiety levels, as measured by the PHQ-4, compared to the control group. The findings of the study suggest that AAT fosters a healing environment for patients, incorporating a holistic and humanistic approach that elicits overwhelmingly positive responses.

Discussion

The outcomes of this meta-analysis provide the long-standing belief that animals can play a beneficial role in the healing process. The study revealed positive and moderately strong results across various aspects, including medical well-being,

behavioral outcomes, and the reduction of autism spectrum symptoms. Moreover, the effect on all four outcomes, which include psychiatric disorders, neurological disorders, stress and anxiety, and depression, were consistent and uniform. Additionally, support for AAT was evident from 4 studies comparing it with established interventions, demonstrating that AAT was equally or more effective. These compelling findings indicate that AAT is a robust intervention deserving of further exploration and use. This systematic review and meta-analysis specifically focused on dogs as the assisting animals in a health care setting. However, there were no limitations on the characteristics of the population included in the study. Although this research synthesis provides evidence in favor of the effectiveness of AAT, it is essential to acknowledge the complexities associated with interventions in general and the specific nuances related to the use of AAT.

The majority of articles included in this systematic review were based on randomized controlled trials conducted in various countries. Additionally, time series and daily announcements, divided according to different conditions, were also considered. The increased number of studies provided greater support in assessing the variance of heterogeneity and potential group differences. Although the results are speculative, the meta-analysis demonstrated homogeneity in the summary values, with only one exploratory group difference reaching statistical significance. Nonetheless, this analysis brought forth several intriguing questions and patterns, serving as a foundation for discussions or further research on the factors influencing the effectiveness of AAT. For instance, consistent benefits were observed in children, young age groups, and old age groups across all outcome variables, including symptoms associated with psychiatric disorders, stress, and anxiety [35]. In particular, among the adult group, a high prevalence of psychiatric disorders, followed by neurological disorders, stress, and anxiety, was found. In contrast, in children, a high number of cases related to stress and anxiety disorders were identified.

Several organizations in different countries are actively working to promote AAT. At the global level, the International Association of Human-Animal Interaction is the worldwide consortium of organizations involved in the practice, research, or education of AAT and the training of service animals [36]. In the United States, the Society for Healthcare Epidemiology of America has established comprehensive guidelines for animals in health care facilities, which emphasize the importance of written policies, designated AAI visit liaisons, and formal training programs for animals and handlers [37]. However, despite these guidelines, there is no legal requirement for health care facilities to adopt these measures. One notable organization in the United States, Pet Partners, stands out as the only national therapy animal organization that mandates volunteer training and biennial evaluations of animal-handler teams, and prohibits raw meat diets [38]. In Europe, the European Society for Animal Assisted Therapy (ESAAT) plays a substantial role as an influential organization operating across various disciplines and professions within the field. ESAAT's primary mission is to accredit education and training programs in the domain of animal-assisted interventions [39]. While the Western world has made significant advancements in AAT, Eastern countries such as India, China, Taiwan, Japan, and Sri Lanka are still in the early stages of exploring and implementing such practices. These countries are currently in the infancy phase of using and developing their own AAT programs. As awareness and understanding of the benefits of AAT continue to grow worldwide, it is expected that these Eastern countries will gradually catch up and further enhance their ATT initiatives [40].

Our review was based on a limited number of studies, which can be attributed to our strict inclusion criteria and the presence of suboptimal study designs. Specifically, many of the randomized trials were characterized by small sample sizes, short durations, and a lack of follow-up assessments. Another limitation pertains to the suitability of the outcome measures used, which may not fully capture the important values and impacts as perceived by the participants. On the other hand, the qualitative research included in the review exhibited higher overall quality and contributed valuable insights to our findings.

In conclusion, the reviewed studies provide preliminary evidence of the potential benefits of AAT in certain conditions. It suggests that dog-assisted therapy can have minor to moderate effects in treating psychiatric disorders, cognitive disorders, neurological disorders, etc, and demonstrates potential in various medical interventions. However, it is important to note that some of the outcome measures analyzed did not show significant effects, and further research is needed to better understand the specific contexts and conditions. To foster the growth of such therapy, we need education campaigns, research programs, professional support, and media awareness to increase the effectiveness of AAT across different countries.

Acknowledgments

This research was funded by VtR Inc-CGU (SCRPD1L0221), DOXABIO-CGU (SCRPD1K0131), and a CGU grant (UZRPD1L0011, UZRPD1M0081).

Conflicts of Interest

None declared.

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist
[DOCX File, 31 KB - xmed_v5i1e51787_app1.docx]

References

1. Thalmann O, Shapiro B, Cui P, et al. Mitochondrial Genomes of ancient Canids suggest a European origin of domestic dogs. Science 2013 Nov 15;342(6160):871-874. [doi: 10.1126/science.1243650] [Medline: 24233726]
2. Gee NR, Rodriguez KE, Fine AH, Trammell JP. Dogs supporting human health and well-being: a biopsychosocial approach. Front Vet Sci 2021 Mar;8:630465. [doi: 10.3389/fvets.2021.630465] [Medline: 33860004]
3. Sahebalzamani M, Rezaei O, Moghadam LF. Animal-assisted therapy on happiness and life quality of chronic psychiatric patients living in psychiatric residential care homes: a randomized controlled study. BMC Psychiatry 2020 Dec 1;20(1):575. [doi: 10.1186/s12888-020-02980-8] [Medline: 33261578]
4. O'Haire ME, Guérin NA, Kirkham AC. Animal-assisted intervention for trauma: a systematic literature review. Front Psychol 2015 Aug 7;6:1121. [doi: 10.3389/fpsyg.2015.01121] [Medline: 26300817]
5. Nurenberg JR, Schleifer SJ, Shaffer TM, et al. Animal-assisted therapy with chronic psychiatric inpatients: equine-assisted psychotherapy and aggressive behavior. Psychiatr Serv 2015 Jan 1;66(1):80-86. [doi: 10.1176/appi.ps.201300524] [Medline: 25269512]
6. Şahin S, Kose B, Zarif M. Animal-assisted therapy in occupational therapy. Occup Ther Therapeutic Creative Use Activity 2018 Nov 5. [doi: 10.5772/intechopen.76468]
7. The IAHAIO Defintions for animal assisted intervention and guidelines for wellness of animals involved in AAI. International Association of Human-Animal Interaction Organizations. 2014. URL: https://iahaio.org/wp/wp-content/uploads/2018/04/iahaio_wp_updated-2018-final.pdf [accessed 2024-03-08]
8. Menna LF, Santaniello A, Amato A, et al. Changes of oxytocin and serotonin values in dialysis patients after animal assisted activities (AAAs) with a dog—a preliminary study. Animals 2019 Aug 3;9:526. [doi: 10.3390/ani9080526]
9. Menna LF, Santaniello A, Gerardi F, et al. Efficacy of animal-assisted therapy adapted to reality orientation therapy: measurement of salivary cortisol. Psychogeriatrics 2019 Sep;19(5):510-512. [doi: 10.1111/psyg.12418] [Medline: 30740833]
10. Menna LF, Santaniello A, Todisco M, et al. The human-animal relationship as the focus of animal-assisted interventions: a One Health approach. Int J Environ Res Public Health 2019 Sep 29;16(19):3660. [doi: 10.3390/ijerph16193660] [Medline: 31569460]
11. Hoagwood KE, Acri M, Morrissey M, Peth-Pierce R. Animal-assisted therapies for youth with or at risk for mental health problems: a systematic review. Appl Dev Sci 2017 Jan;21(1):1-13. [doi: 10.1080/10888691.2015.1134267] [Medline: 28798541]
12. Muldoon AL, Kuhns LM, Supple J, Jacobson KC, Garofalo R. A web-based study of dog ownership and depression among people living with HIV. JMIR Ment Health 2017 Nov 8;4(4):e53. [doi: 10.2196/mental.8180] [Medline: 29117933]
13. Albuquerque N, Guo K, Wilkinson A, Savalli C, Otta E, Mills D. Dogs recognize dog and human emotions. Biol Lett 2016 Jan;12(1):20150883. [doi: 10.1098/rsbl.2015.0883] [Medline: 26763220]
14. Hare B, Brown M, Williamson C, Tomasello M. The domestication of social cognition in dogs. Science 2002 Nov 22;298(5598):1634-1636. [doi: 10.1126/science.1072702] [Medline: 12446914]
15. Miklósi A, Kubinyi E, Topál J, Gácsi M, Virányi Z, Csányi V. A simple reason for a big difference: wolves do not look back at humans, but dogs do. Curr Biol 2003 Apr 29;13(9):763-766. [doi: 10.1016/s0960-9822(03)00263-x] [Medline: 12725735]
16. Payne E, Bennett PC, McGreevy PD. Current perspectives on attachment and bonding in the dog–human dyad. Psychol Res Behav Manag 2015 Feb 24;8:71-79. [doi: 10.2147/PRBM.S74972] [Medline: 25750549]
17. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Training. URL: https://training.cochrane.org/handbook [accessed 2023-06-24]
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Thompson Coon J, Gwernan-Jones R, Garside R, et al. Developing methods for the overarching synthesis of quantitative and qualitative evidence: the interweave synthesis approach. Res Synth Methods 2020 Jul;11(4):507-521. [doi: 10.1002/jrsm.1383] [Medline: 31725951]
20. Shih CA, Yang MH. Effect of animal-assisted therapy (AAT) on social interaction and quality of life in patients with schizophrenia during the COVID-19 pandemic: an experimental study. Asian Nurs Res (Korean Soc Nurs Sci) 2023 Feb;17(1):37-43. [doi: 10.1016/j.anr.2023.01.002] [Medline: 36646276]
21. Chen CR, Hung CF, Lee YW, Tseng WT, Chen ML, Chen TT. Functional outcomes in a randomized controlled trial of animal-assisted therapy on middle-aged and older adults with schizophrenia. Int J Environ Res Public Health 2022 May 21;19(10):6270. [doi: 10.3390/ijerph19106270] [Medline: 35627807]
22. Allen B, Shenk CE, Dreschel NE, et al. Integrating animal-assisted therapy into TF-CBT for abused youth with PTSD: a randomized controlled feasibility trial. Child Maltreat 2022 Aug;27(3):466-477. [doi: 10.1177/1077559520988790] [Medline: 33499659]
23. Chen TT, Hsieh TL, Chen ML, Tseng WT, Hung CF, Chen CR. Animal-assisted therapy in middle-aged and older patients with schizophrenia: a randomized controlled trial. Front Psychiatry 2021 Aug 3;12:713623. [doi: 10.3389/fpsyt.2021.713623] [Medline: 34456769]

24. Anderson D, Brown S. The effect of animal-assisted therapy on nursing student anxiety: a randomized control study. Nurse Educ Pract 2021 Mar;52:103042. [doi: 10.1016/j.nepr.2021.103042] [Medline: 33839594]
25. Thakkar TK, Naik SN, Dixit UB. Assessment of dental anxiety in children between 5 and 10 years of age in the presence of a therapy dog: a randomized controlled clinical study. Eur Arch Paediatr Dent 2021 Jun;22(3):459-467. [doi: 10.1007/s40368-020-00583-1] [Medline: 33245525]
26. Santaniello A, Garzillo S, Amato A, et al. Animal-assisted therapy as a non-pharmacological approach in Alzheimer's disease: a retrospective study. Animals (Basel) 2020 Jul 6;10(7):1-10. [doi: 10.3390/ani10071142] [Medline: 32640648]
27. Brown S, Snelders J, Godbold J, et al. Effects of animal-assisted activity on mood states and feelings in a psychiatric setting. J Am Psychiatr Nurses Assoc Nov/Dec 2020;26(6):555-567. [doi: 10.1177/1078390319853617] [Medline: 31179871]
28. Hinic K, Kowalski MO, Holtzman K, Mobus K. The effect of a pet therapy and comparison intervention on anxiety in hospitalized children. J Pediatr Nurs 2019 ;46:55-61. [doi: 10.1016/j.pedn.2019.03.003] [Medline: 30852256]
29. Ginex P, Montefusco M, Zecco G, et al. Animal-facilitated therapy program: outcomes from caring canines, a program for patients and staff on an inpatient surgical oncology unit. Clin J Oncol Nurs 2018 Apr 1;22(2):193-198. [doi: 10.1188/18.CJON.193-198] [Medline: 29547605]
30. MB P. Effectiveness of animal assisted therapy on children and adolescents with autism. Int J Res Soc Sci 2018 Feb;8(2):2249-2496 [FREE Full text]
31. McCullough A, Ruehrdanz A, Jenkins MA, et al. Measuring the effects of an animal-assisted intervention for pediatric oncology patients and their parents: a multisite randomized controlled trial. J Pediatr Oncol Nurs 2018 May;35(3):159-177. [doi: 10.1177/1043454217748586] [Medline: 29268667]
32. Branson SM, Boss L, Padhye NS, Trötscher T, Ward A. Effects of animal-assisted activities on biobehavioral stress responses in hospitalized children: a randomized controlled study. J Pediatr Nurs 2017 ;36:84-91. [doi: 10.1016/j.pedn.2017.05.006]
33. Stefanini MC, Martino A, Allori P, Galeotti F, Tani F. The use of animal-assisted therapy in adolescents with acute mental disorders: a randomized controlled study. Complement Ther Clin Pract 2015 Feb;21(1):42-46. [doi: 10.1016/j.ctcp.2015.01.001] [Medline: 25701449]
34. Menna LF, Santaniello A, Gerardi F, Di Maggio A, Milan G. Evaluation of the efficacy of animal-assisted therapy based on the reality orientation therapy protocol in Alzheimer's disease patients: a pilot study. Psychogeriatrics 2016 Jul;16(4):240-246. [doi: 10.1111/psyg.12145] [Medline: 26370064]
35. Chiu CJ, Hsieh S, Li CW. Needs and preferences of middle-aged and older adults in Taiwan for companion robots and pets: survey study. J Med Internet Res 2021 Jun 11;23(6):e23471. [doi: 10.2196/23471] [Medline: 34347621]
36. Who we are. International Association of Human-Animal Interaction Organizations. URL: https://iahaio.org/missions-goals/ [accessed 2023-06-30]
37. Murthy R, Bearman G, Brown S, et al. Animals in healthcare facilities: recommendations to minimize potential risks. Infect Control Hosp Epidemiol 2015 May;36(5):495-516. [doi: 10.1017/ice.2015.15] [Medline: 25998315]
38. Pet Partners. URL: https://petpartners.org/ [accessed 2023-06-28]
39. ESAAT. About us. URL: https://www.esaat.org/en/about-us/ [accessed 2023-06-29]
40. Narvekar HN. A reflection on the current status of animal-assisted therapy in India. Hum Arenas 2021 Sep 17;6(2023):760-775. [doi: 10.1007/s42087-021-00250-x]

Abbreviations

AAT: animal-assisted therapy
ESAAT: European Society for Animal Assisted Therapy
GDS: Geriatric Depression Scale
MeSH: Medical Subject Headings
MMSE: Mini-Mental State Examination
PHQ-4: Patient Health Questionnaire–4
PICOS: Patient, Problem, or Population; Intervention; Comparison; Outcomes; and Study Design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
STAI: State-Trait Anxiety Inventory

Edited by E Meinert, G Eysenbach; submitted 11.08.23; peer-reviewed by Anonymous, Anonymous; revised version received 16.12.23; accepted 27.12.23; published 18.03.24.

Please cite as:
Pandey RP, Himanshu , Gunjan , Mukherjee R, Chang CM
The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence
JMIRx Med 2024;5:e51787
URL: https://xmed.jmir.org/2024/1/e51787
doi:10.2196/51787

Correction: Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018

Ali Hazim Mustafa[1,*], MBChB, HD CM; Hanan Abdulghafoor Khaleel[2,*], MBChB, MSc CM, DrPH; Faris Lami[3,*], MBChB, PhD

[1]

[2]

[3]

[*]all authors contributed equally

Corresponding Author:
Hanan Abdulghafoor Khaleel, MBChB, MSc CM, DrPH

Related Article:

Companion article: https://med.jmirx.org/2024/1/e54611

(JMIRx Med 2024;5:e64958) doi:10.2196/64958

In “Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018” (JMIRx Med 2024;5:e54611) the authors noted one error.

The originally published manuscript was missing some figures. The following figures have been added to the published paper (Figures 4-10):

Figure 1. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2013.

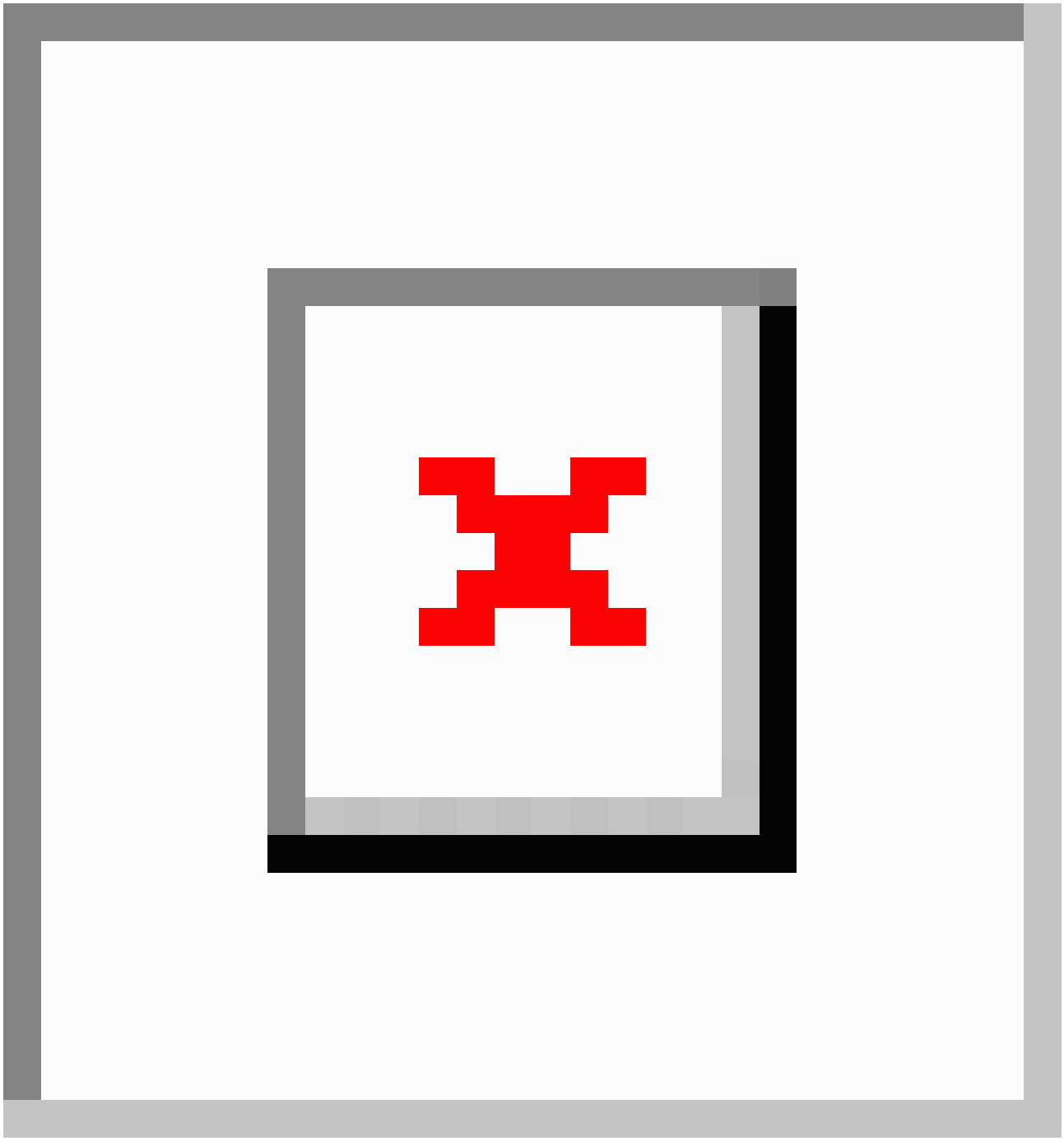


Figure 2. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2014.

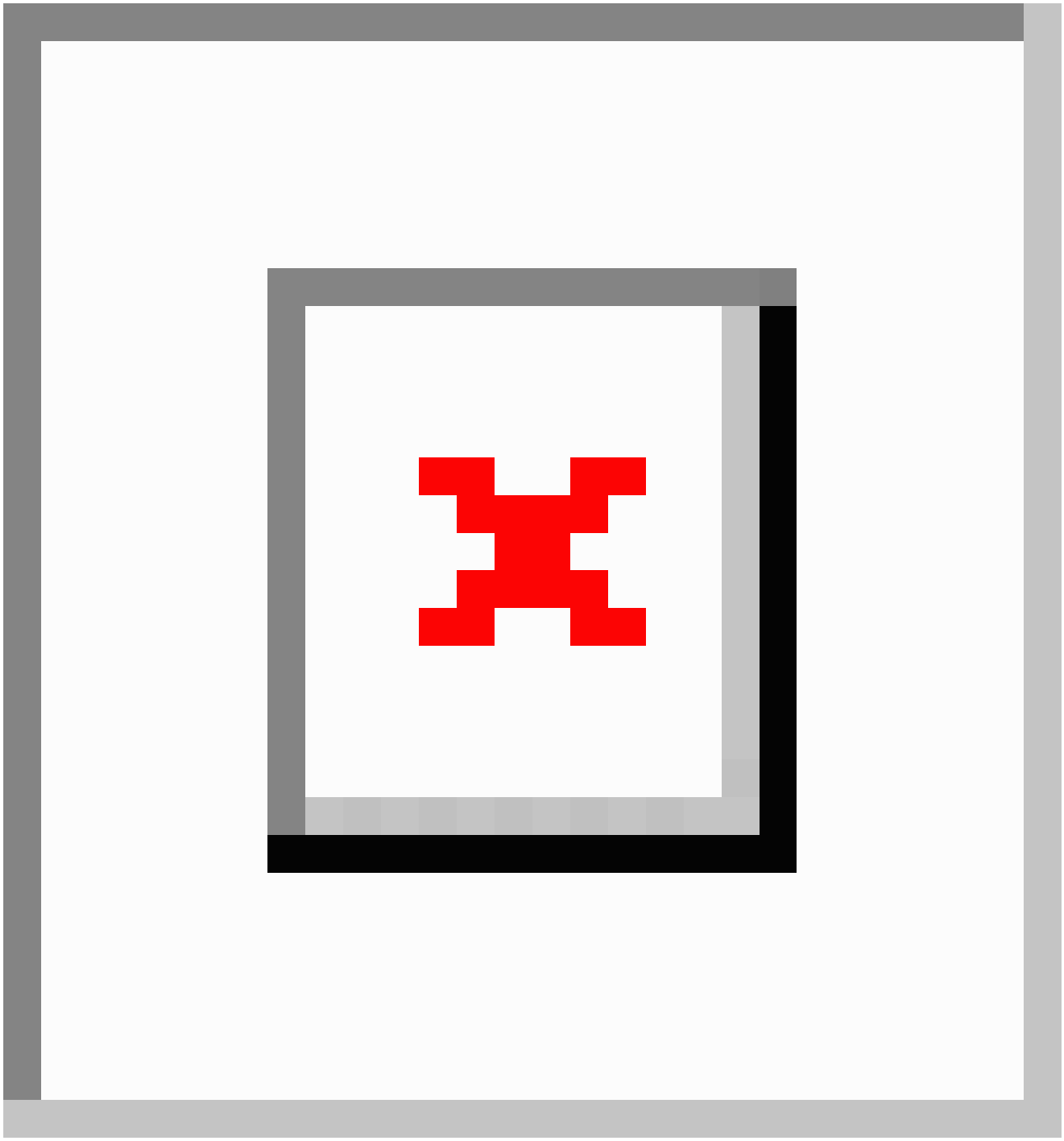


Figure 3. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2015.

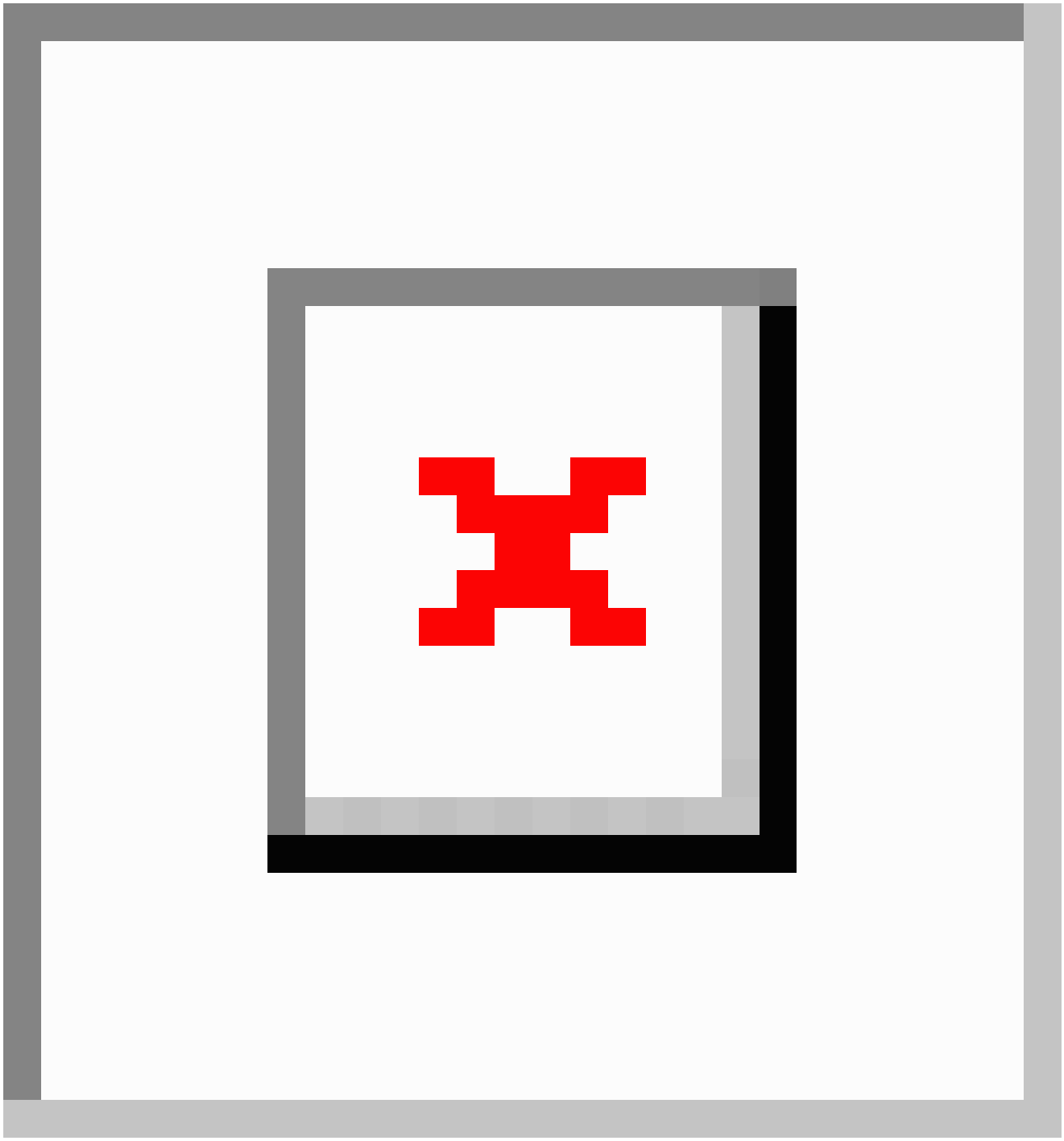


Figure 4. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2016.

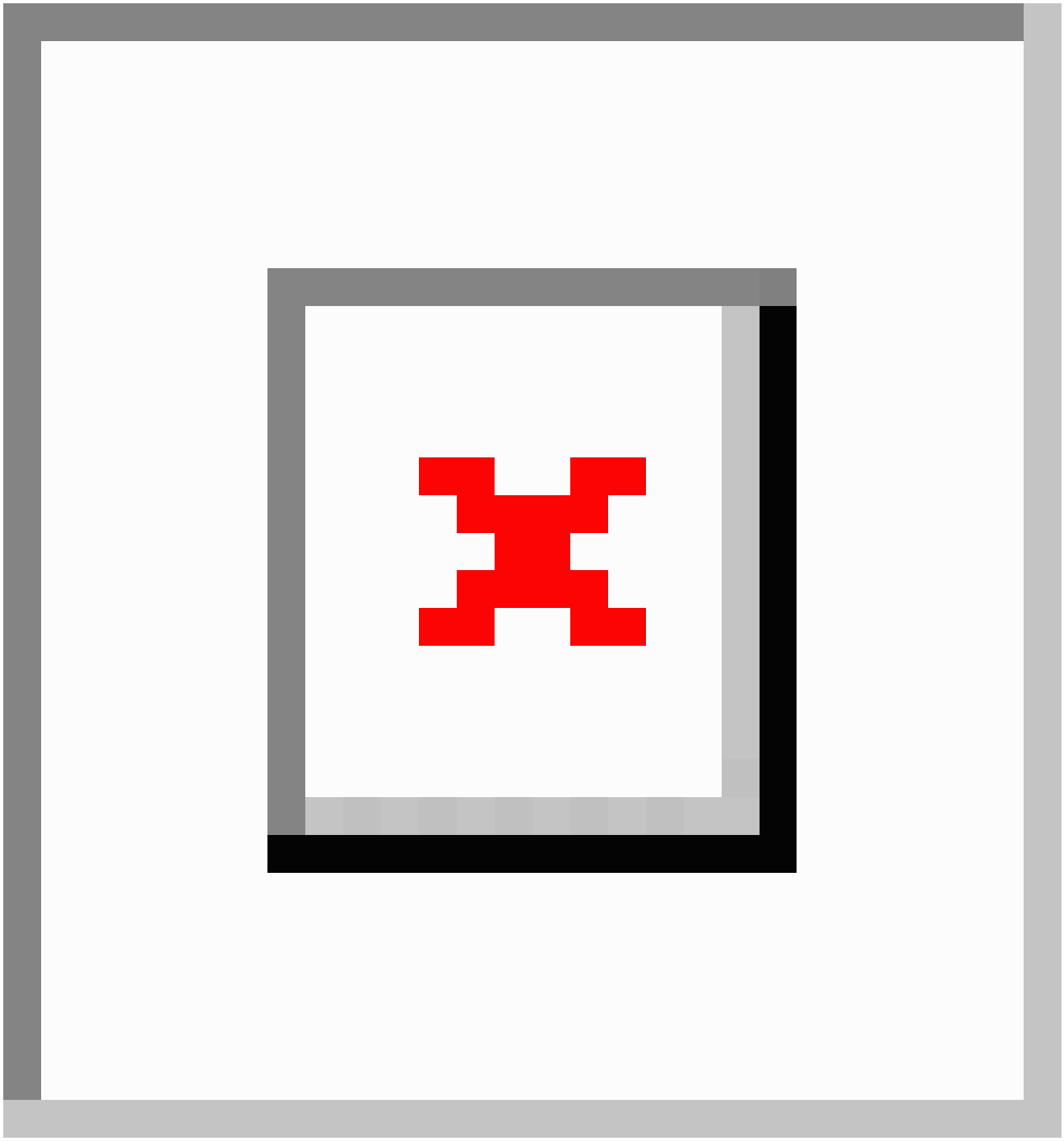


Figure 5. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2017.

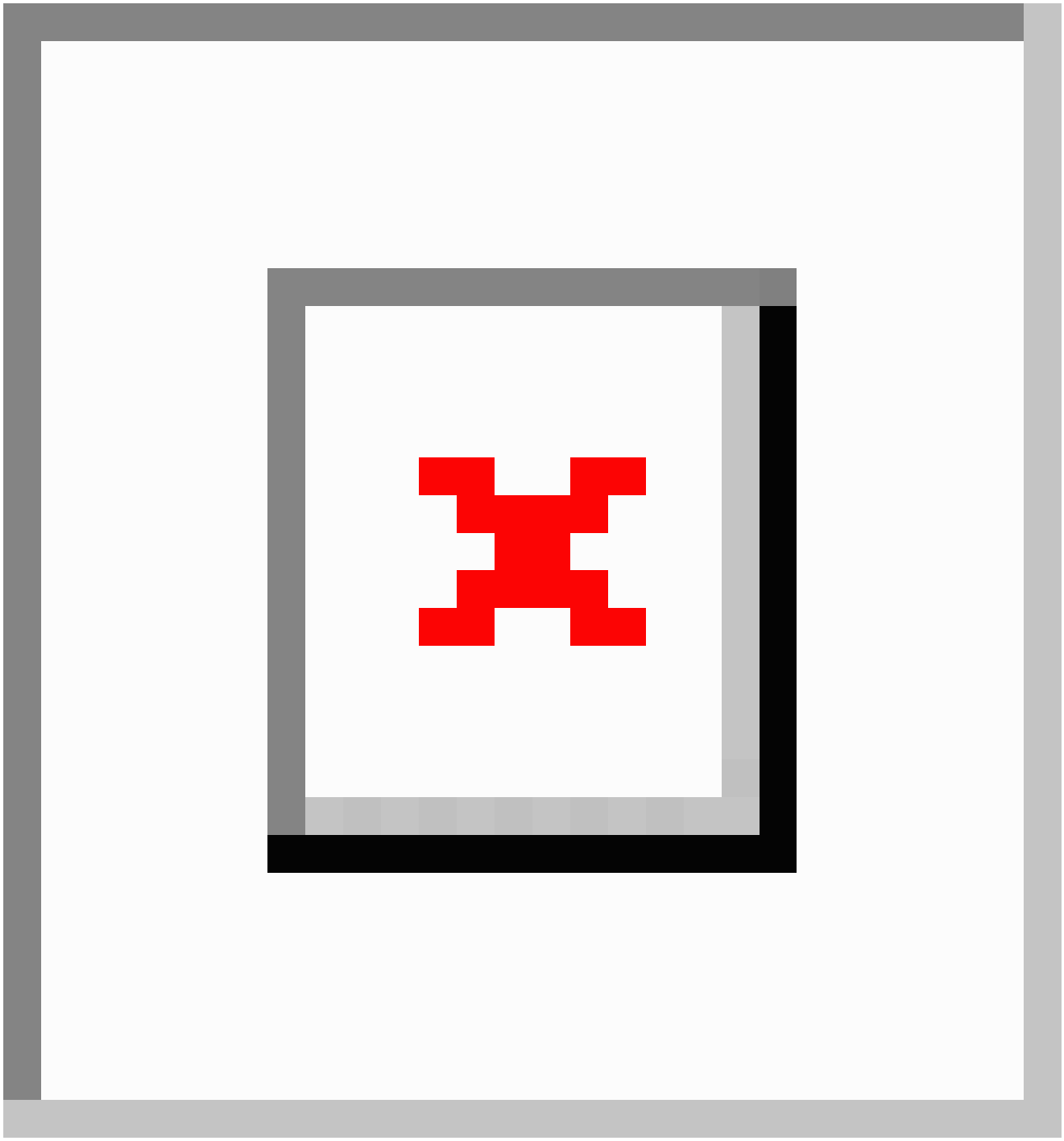


Figure 6. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2018.

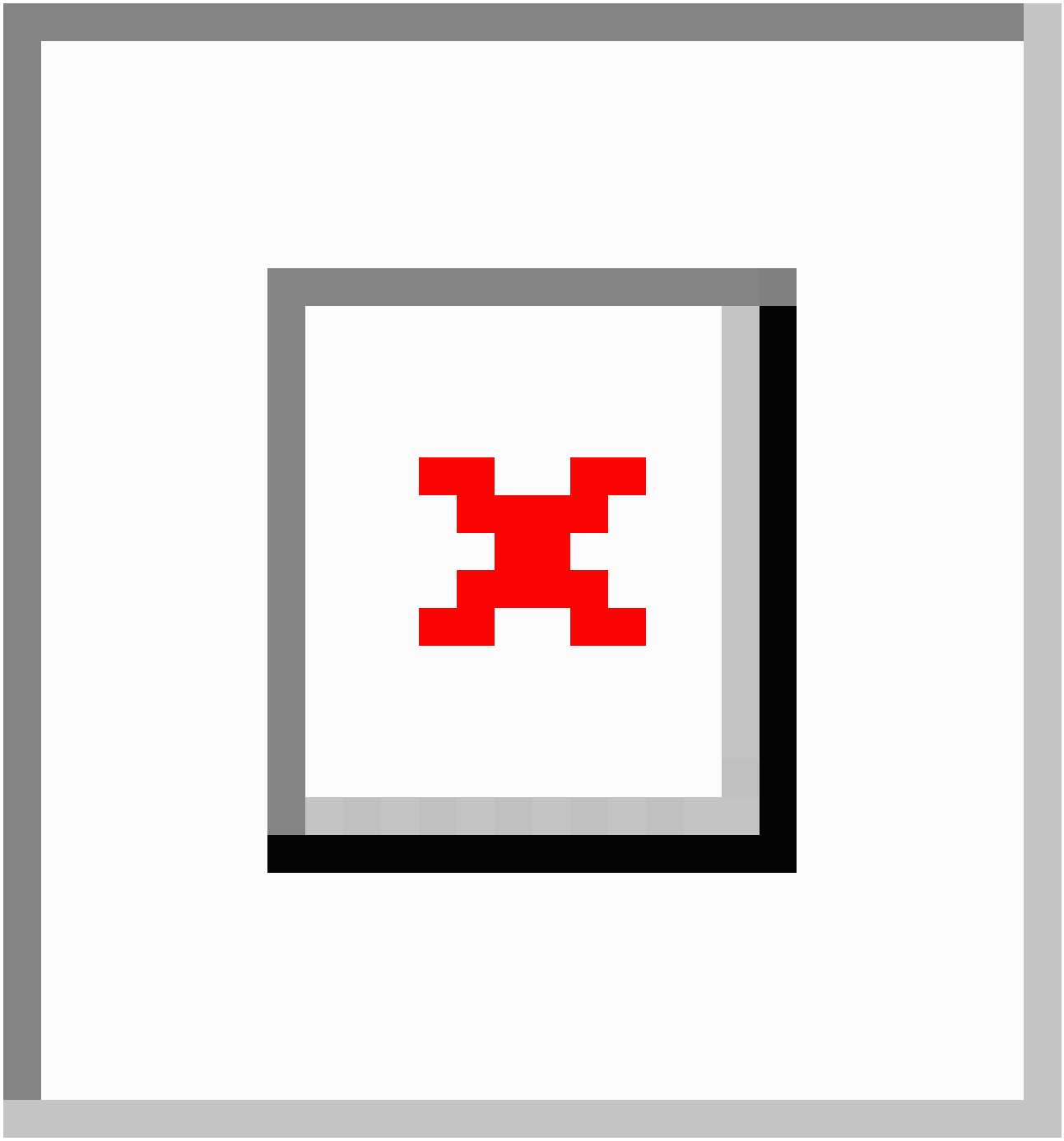
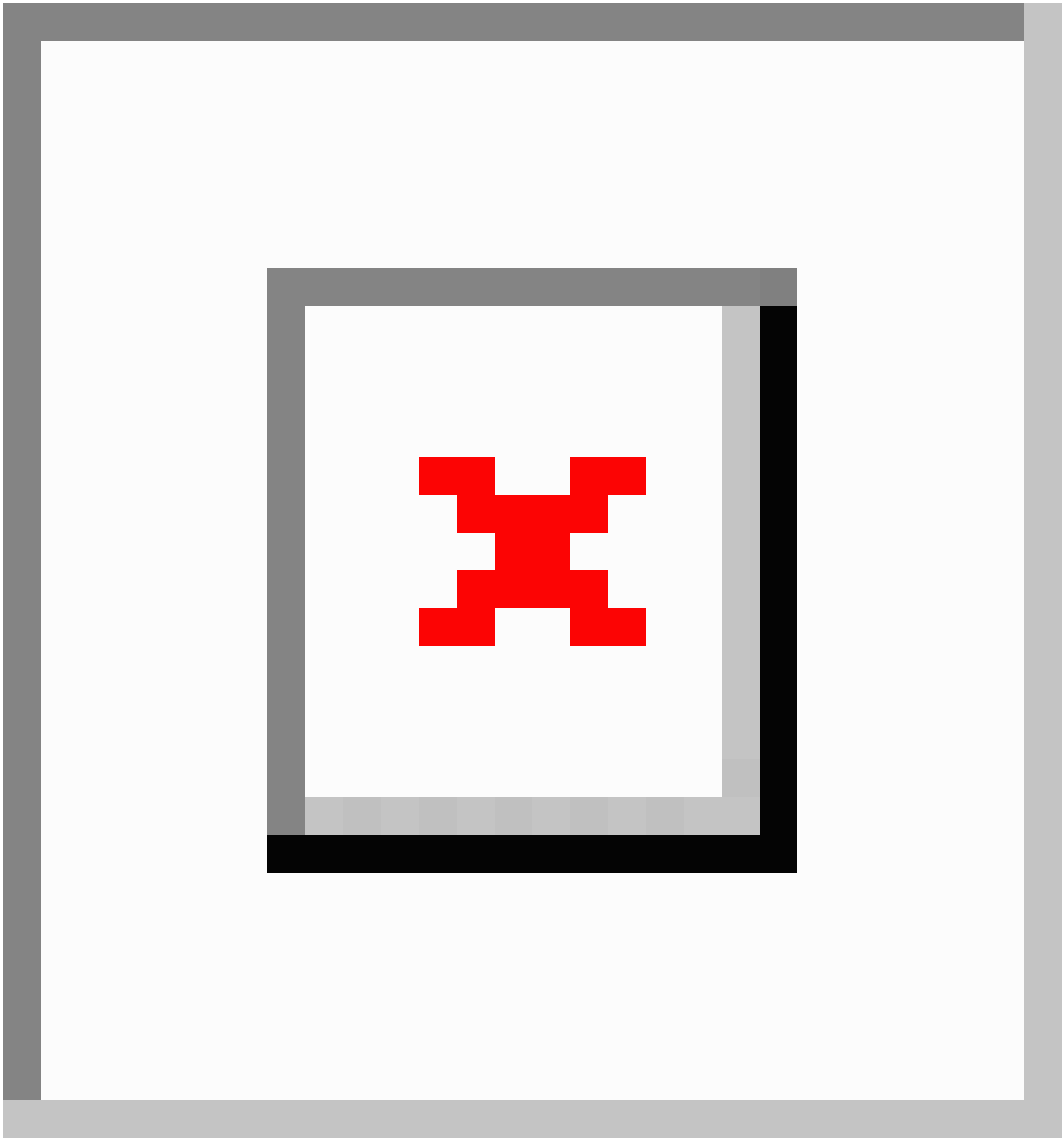


Figure 7. Spatial distribution maps of human brucellosis incidence rate in 2013 (A) and 2018 (B).



In addition, an in-text reference to these figures has been added to the final sentence of the Results section:

> *District Koisanjaq in Erbil shifted from a hot spot in 2014 to a cold spot in 2015 (Figures 4-10).*

The correction will appear in the online version of the paper on the JMIR Publications website on August 9, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Conflicts of Interest

None declared.

Submitted 31.07.24; this is a non–peer-reviewed article;accepted 31.07.24; published 09.08.24.

Please cite as:
Mustafa AH, Khaleel HA, Lami F
Correction: Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018
JMIRx Med 2024;5:e64958
URL: https://xmed.jmir.org/2024/1/e64958
doi:10.2196/64958

Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45688

Companion article: https://med.jmirx.org/2024/1/e55863

Companion article: https://med.jmirx.org/2024/1/e45688

(JMIRx Med 2024;5:e55900) doi:10.2196/55900

KEYWORDS

COVID-19; injury; prevalence; adaptation; sports medicine; follow-up; training; football; epidemiology; sport; athlete; athletic; injuries

This is the peer-review report for "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study."

Round 1 Review

This paper, "COVID-19 NFL Injury Prevalence Analysis, A Follow-Up Study" [1], is an interesting read. The conclusions drawn in the paper are supported by the data. All the related works are cited appropriately. The limitations of the presented study are discussed appropriately.

My only concern is about the 2 histograms. The authors should make these histograms more clear, readable, and engaging. Use standard deviation or error bars wherever applicable.

Conflicts of Interest

None declared.

Reference

1. Puga TB, Schafer J, Thiel G, et al. COVID-19 National Football League (NFL) injury analysis: follow-up study. JMIRx Med 2024;5:e45688. [doi: 10.2196/45688]

Edited by E Meinert; submitted 28.12.23; this is a non–peer-reviewed article;accepted 28.12.23; published 13.02.24.

Please cite as:
Anonymous
Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"
JMIRx Med 2024;5:e55900
URL: https://xmed.jmir.org/2024/1/e55900
doi:10.2196/55900

Peer Review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers"

Anna Marin, PhD

Icahn School of Medicine at Mount Sinai, New York, NY, United States

Related Articles:

Companion article: http://preprints.jmir.org/preprint/42211

Companion article: https://www.medrxiv.org/content/10.1101/2022.05.25.22275517v2

Companion article: https://med.jmirx.org/2024/1/e56439

Companion article: https://med.jmirx.org/2024/1/e42211

(JMIRx Med 2024;5:e56443) doi:10.2196/56443

KEYWORDS

Latino; dementia; caregiving; COVID-19; Alzheimer's disease; health disparity; qualitative research; transcript analysis; health inequality; minority population; epidemiology; peer support; social services; health care; Alzheimer's; minority; qualitative; interview; caregiver; primary care; impact; resilience; disparity; outcome; Alzheimer disease; Alzheimer

This is the peer-review report for "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers."

Round 1 Review

General Comments

This paper [1] provides an in-depth qualitative assessment of the impact and resilience factors related to the COVID-19 pandemic among Latino families with Alzheimer disease and related dementias (ADRD). The authors interviewed 21 family caregivers and 23 primary care providers (PCPs) across the United States and identified 2 primary themes that characterized the experiences of the participants, involving both the impact of the pandemic and the strategies adopted to cope with the detrimental impact of the pandemic. The topic covered is of significant importance and provides important background to better understand the impact of the COVID-19 pandemic on Latino families with ADRD, with the aim of improving the quality of care and equity among the Latino community. Overall, I think that the authors should reorganize the results to better align with the aim of the study. In my specific comments I have included specific suggestions on how to make the results more organized and succinct. To strengthen the generalizability and interpretation of the findings, the authors should also include a descriptive quantitative analysis of the interviews' analysis, where the reader can examine the prevalence of each theme and subtheme among the 21 family members and 23 PCPs.

Specific Comments

Major Comments

1. The results section of the abstract is not very clear. What are the overall findings? How have the 2 themes been identified?

2. Introduction: Please outline the qualitative variables selected to investigate the aim of the study.

3. Methods:

a. Sample and assessment: Given the qualitative nature of the study, it would be helpful if the authors included an example of a question from the interview script and also attached as an appendix the template of questions they used to direct the conversation.
b. Data analysis: This section is not very clear, and it would help if the authors could describe the process of coding in a step-by-step manner and also make the text more succinct.

4. Results: The authors provide a very detailed qualitative analysis; however, more quantitative information should be provided to show the prevalence of each theme and subtheme for all the caregivers and PCPs (eg, how many PCPs reported food access and malnutrition during the COVID-19 pandemic?)

Minor Comments

5. In the abstract, results section, please remove the capital letters for the 2 themes and consider writing them as "Qualitative analysis of transcripts revealed two themes: (1) the impact of a global pandemic (eg, accelerated cognitive and physical decline, or caregivers choosing between risking finances and the family's

infection given the work situation) and (2) developing resilience to the effects of the pandemic (eg, caregivers seeking vaccination sites, moving in with the care recipient and adopting telehealth."

6. Introduction: This sentence should be revised for clarity: "As of January of 2022, Latinos represent 8% of the US 65 and older population, but 13% of COVID-19 cases in the same age group [9]." It is not clear if the 13% of cases is the percentage of COVID-19 cases in Latino individuals aged 65 years and older.

7. Methods:

a. This sentence should be incorporated in the introduction or removed: "The goal of this study was to gain an in-depth understanding of the impact of the COVID-19 pandemic on Latino families with ADRD."
b. Who transcribed the interviews?

8. Results:

a. To improve clarity, the authors should label the descriptions of the different themes as themes (1, 2, etc) and subthemes (1.1, 1.2, 1.3, etc; 2.1, 2.2, 2.3, etc). Please add this labeling both in Table 2 and in the text below to help the reader better orient into each of the themes.

9. Page 6 last sentence: Please remove "make" from "To make bring rigor and validity."

10. Page 14, last paragraph: The authors should edit "work" with "works."

Round 2 Review

General Comments

This paper provides an in-depth qualitative assessment of the impact and resilience factors related to the COVID-19 pandemic among Latino families with ADRD. The authors interviewed 21 family caregivers and 23 PCPs across the United States. They identified 2 primary themes that characterized the participants' experiences, involving both the impact of the pandemic and the strategies adopted to cope with the detrimental impact of the pandemic. The topic covered is of significant importance and provides important background to better understand the impact of the COVID-19 pandemic on Latino families with ADRD and to improve the quality of care and equity among the Latino community. The authors did a great job improving the clarity of the methods and the results. I have included other minor revisions to further improve the clarity of the text.

Specific Comments

Minor Comments

1. Results:
   a. Paragraph "theme 8" (line 4): Remove "and" before "healthcare."
   b. Paragraph 8.3 (line 3): Replace "requested" with "requesting."
2. Discussion: Please clarify the second paragraph of the "implication and future directions" section. In the first sentence, "Our findings can inform future studies. For example, participants reported pandemic-related physical and cognitive deterioration and the importance of family support." Can you clarify what type of findings can inform future research? Could you say something on the line of "Our findings regarding the physical and cognitive deterioration caused by the pandemic and the importance of family support may help inform future studies on…" In the same paragraph, there is a very minor typo; please replace "health" with "healthy."

Conflicts of Interest

None declared.

Reference

1. Perales-Puchalt J, Peltzer J, Fracachan-Cabrera M, et al. Impact of the COVID-19 pandemic on Latino families with Alzheimer disease and related dementias: qualitative interviews with family caregivers and primary care providers. JMIRx Med 2024;5:e42211. [doi: 10.2196/42211]

Abbreviations

ADRD: Alzheimer disease and related dementias
PCP: primary care provider

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 08.03.24.

Please cite as:
Marin A
Peer Review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers"
JMIRx Med 2024;5:e56443
URL: https://xmed.jmir.org/2024/1/e56443
doi:10.2196/56443

Peer Review of “Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study”

Lucía Carrasco-Ribelles, BEng, MSc

Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Barcelona, Spain

Related Articles:

Companion article: http://preprints.jmir.org/preprint/50803

Companion article: https://med.jmirx.org/2024/1/e56441

Companion article: https://med.jmirx.org/2024/1/e50803

(JMIRx Med 2024;5:e56446) doi:10.2196/56446

KEYWORDS

artificial intelligence; adoption; acceptance; opinion; perceptions; survey; expectations; physician; medical survey; qualitative study; AI

This is the peer-review report for “Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study.”

Round 1 Review

General Comments

This paper [1] reports the results of a survey on medical expectations on artificial intelligence (AI) solutions in daily practice. The authors argue that it is important to know the opinion that physicians would have as users of these solutions, and the reviewer could not agree more. Therefore, the results of this work may be of interest to the community.

Specific Comments

Major Comments

1. The authors say that these results represent the opinion of Brazilian physicians. Perhaps that is a bit presumptuous, at least without somehow justifying the size of the hospital relative to the Brazilian population. What percentage of the Brazilian population attends this hospital? What percentage of Brazilian physicians works there?

2. I have not been able to find the supplementary material anywhere. Therefore, I could not review the complete questionnaire.

3. The division into <20 years of practice and >20 years of practice does not seem sufficient to this reviewer, since in <20 years of practice you can still have quite senior physicians. I would add an additional division: <10 years, 10-20 years, and >20 years of practice.

Minor Comments

4. How are the percentages calculated in Table 1? The percentages of every column should sum up to 100.

5. Could the authors comment on, if the physicians reported it in the questionnaire, which AI solutions they used in their daily life? Are they used in their personal life or in their work?

6. I assume there is an issue with the color legend for “Work facilitation” in Figure 2.

7. I would not only say that physicians think AI will not interfere with the number of appointments. A third of them thinks that AI solutions will increase the number of appointments.

8. I would include, if possible, a subanalysis of the responses per gender and discuss if there are any differences.

Round 2 Review

General Comments

This reviewer thanks the authors for the work done to improve the quality of the paper with this revision. However, I still have some comments.

Specific Comments

Major Comments

1. In the previous review round, I asked about the AI solutions the health care workers used in their daily life. The authors replied by saying “the specific app (which uses AI algorithms) in their daily lives was not asked, but we believe it is the same as most of the people in Brazil: Instagram, WhatsApp, Waze, Google Apps, Alexa, Siri, Twitter and banks app.” This reviewer thinks this should be commented somewhere in the manuscript. From this questionnaire question, it seemed that workers have access to true AI solutions in their daily lives. However, these apps the authors mentioned as “AI solutions” use AI in their workflow but are not entirely based on AI and should not be considered “AI solutions.” Without commenting on this, the reader may think that the experience of this population in the use of AI is greater than it really is.

Minor Comments

2. I have not yet been able to access the supplementary material, and the color legend in Figure 2 is still not fixed.

3. In the text, it appears as *P*=.079, which is not significant. Please check.

4. The *P*=.0513 in Table 2 is not significant.

5. There should be a “Total” column in Table 1.

Conflicts of Interest

None declared.

Reference

1. Giavina-Bianchi M, Amaro Jr E, Machado BS. Medical expectations of physicians on AI solutions in daily practice: cross-sectional survey study. JMIRx Med 2024;5:e50803. [doi: 10.2196/50803]

Abbreviations

AI: artificial intelligence

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 25.03.24.

Please cite as:
Carrasco-Ribelles L
Peer Review of “Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study”
JMIRx Med 2024;5:e56446
URL: https://xmed.jmir.org/2024/1/e56446
doi:10.2196/56446

Peer Review of "Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis"

Shanmukha Gorthy, BDS, MDS

Drs Sudha & Nageswara Rao Siddhartha Institute of Dental Sciences, Chinoutpalli, India

Related Articles:

Companion article: https://preprints.jmir.org/preprint/56759

Companion article: https://www.medrxiv.org/content/10.1101/2024.01.11.24301001v1

Companion article: https://med.jmirx.org/2024/1/e56759

(JMIRx Med 2024;5:e60329) doi:10.2196/60329

KEYWORDS

density; teeth; tooth; dental; dentist; dentists; dentistry; oral; tissue; enamel; dentin; Hounsfield; pathology; pathological; radiology; radiological; image; images; imaging; teeth density; Hounsfield unit; diagnostic imaging

This is the peer-review report for "Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis."

Round 1 Review

General Comments

This paper [1] has a very good topic selected by the authors. Based on the study of Hounsfield units in cone-beam computed tomography (CBCT), we can evaluate conditions that are abnormal in patients.

Specific Comments

There is not much to comment on, but try to include good pictures of the CBCT.

Major Comments

1. The authors did not mention the ideal values of the Hounsfield units for enamel and dentin.

2. The authors did not properly explained why there were no gender differences.

3. The authors did not explain why there were the fewest density values for second primary molars and the most density values for maxillary central incisors.

Conflicts of Interest

None declared.

Reference

1. Reshetnikov A, Shaikhattarova N, Mazurok M, Kasatkina N. Dental tissue density in healthy children based on radiological data: retrospective analysis. JMIRx Med 2024;5:e56759. [doi: 10.2196/56759]

Abbreviations

CBCT: cone-beam computed tomography

Edited by E Meinert; submitted 07.05.24; this is a non–peer-reviewed article;accepted 07.05.24; published 20.06.24.

Please cite as:
Gorthy S
Peer Review of "Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis"
JMIRx Med 2024;5:e60329
URL: https://xmed.jmir.org/2024/1/e60329
doi:10.2196/60329

Peer Review of "Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study"

Jamie Sewan Johnston, PhD

Stanford Center for Health Education, Stanford University, Stanford, CA, United States

Related Articles:

Companion article: https://preprints.jmir.org/preprint/48213

Companion article: https://med.jmirx.org/2024/1/e60266

Companion article: https://med.jmirx.org/2024/1/e48213

(JMIRx Med 2024;5:e60430) doi:10.2196/60430

KEYWORDS

primary health care; mother and child health; community health worker; slums; digital applications; health communication

This is the peer-review report for "Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study."

Round 1 Review

General Comments

This study [1] draws on multiple sources of data to assess the efficacy and feasibility of a video-based mobile health (mHealth) tablet intervention used to train and equip 10 community health workers (CHWs) in two slums in Pakistan. The overall strength of the paper is that the authors have collected in-depth qualitative data that can help inform the field on how to build and distribute such an intervention to improve the needs in low-resource communities. The paper should be strengthened by pinpointing the unique and new contributions of the study findings to inform the field on digital health education interventions for CHWs. While the intervention is described in detail, more work is needed to explicate why this study expands our understanding of digital health education for CHWs in deprived settings.

Specific Comments

Major Comments

1. What the authors have found is largely expected and demonstrated in other work globally. Not surprisingly, they find that there is a severe lack of knowledge and critical need for education among CHWs and their patients to bring about behaviors that can improve maternal child health in low-resource communities. It is also not surprising that a well-resourced, highly supported, small-scale pilot (of just 10 CHWs trained) would be successful. While these findings are all important to describe in detail (as the authors have done), to provide more value in this field of work, I suggest the paper more explicitly emphasize what contribution the study adds to the literature. What are the new and important takeaways to improve how mHealth education interventions for CHWs are developed?
2. In addition to more explicitly pointing to the contributions of their findings, the paper could be strengthened considerably by a discussion of how their findings can inform how this small-scale pilot can be taken to scale effectively. The authors allude to this, but I think more could be added with regard to cost-effectiveness. It sounds like an expensive and involved intervention—to my understanding, providing a tablet to CHWs, hosting a two-day training, overseeing an apprentice week, and a refresher training, all on top of development of 14 videos and a calendar for patients. Information on the costs to develop and implement this intervention could be better described, and I would appreciate a critical lens on what would be needed to scale, including identification of barriers. I think the limitations they described with respect to CHW availability, connectivity, etc, should be folded into this discussion. I think this would set up well the ongoing work they describe to test real-world effectiveness in 250 mother-infant pairs.
3. A corollary to the above comment, because the intervention involves so many moving parts (ie, provision of a device, development of videos, in-person training and supervision), do their findings point to particular components of the intervention that are particularly important?
4. The introduction of the paper starts by highlighting the distinctions between the definitions of inequality, disparity, and inequity. I don't think the comparison adds any value to the introduction. In fact, I was confused because, after the first paragraph, implications of an expensive mHealth intervention for equity are not discussed at all. Just because a study is conducted in a low-resource setting, does not

mean that it is working to improve equity. If the authors want to focus on equity, I would appreciate a more critical lens on how the high costs of a digital intervention met with barriers like internet connectivity improve the situation of the poorest communities. (Otherwise, I suggest changing the introduction paragraph.) A video-based tablet intervention that relies upon internet could ostensibly be more effective in communities with more infrastructure and resources, and when scaled more widely in better resourced communities, digital interventions may actually broaden the gap between the haves and the have-nots. How can we think about ways digital interventions can be implemented to ensure this does not happen? (Is this why the in-person CHW-to-patient link is so important? Can this be unpacked?)

5. The phase two findings draw heavily upon the "qualitative feedback" obtained from CHWs and mothers about the Sehat Ghar application and tablet use, but there are scant details on how these data were collected and analyzed. If these qualitative findings are so prominent in the results and not merely anecdotal and complementary to other findings, they need to be described with the level of detail the preintervention in-depth interviews and focus group discussions were described. Were semistructured guides used? What framework was applied to the development of qualitative protocols? How were the data coded and analyzed? How many CHWs and mothers participated?
6. How were the two slums selected for focus? What inclusion/exclusion criteria, if any, were applied when thinking about geographic selection? How do the two slums generalize to the larger set of slums in Islamabad?
7. More of a description of the health systems in the slums is helpful for readers who are not familiar. How do CHWs fit into the larger health system? To my understanding, the 15 CHWs that were included in this study were completely new to the profession as "we identified volunteer women willing to become CHWs." Why focus on completely new volunteers for the study rather than drawing upon the existing CHW workforce? Is it because there were not CHWs already at work in these areas? If there are other CHWs already serving these slums, can this be better described? Please also speak to the generalizability of the findings given that the CHWs in this study started with a much lower lack of knowledge given their novice status. If the study were to be done with experienced CHWs, perhaps the delta in knowledge gains would not be nearly as large.
8. A more detailed description of how participants (health care providers, CHWs, and mothers) were recruited for the study is needed. What was the sampling frame? What were the inclusion/exclusion criteria? What was the consent rate? What roles did the health care providers hold (ie, were they doctors, nurses, other roles)?
9. What were the protocols for conducting observations? Were they unannounced or were CHWs prepared in advance to know that the supervisors would be conducting the observations? Can you address limitations with respect to bias, as individuals generally behave quite differently when observed?
10. Do the authors have any analytics (ie, frequency of video views, engagement with the app) from the tablet/application that can be used to support the observation data and qualitative feedback on the intervention feasibility?

Minor Comments

1. In the Abstract, identify the larger geographic location of the communities.
2. Is there a more recent citation than the 2015 reference used for [2]?
3. The Methods section says that the initial five transcripts were coded independently by two members of the team. What about the remaining transcripts? Were there any checks/reconciliation on the coding of the remaining transcripts?
4. Consider moving some of the details of the intervention, including on page 7 of the Results, to the Methods section. When reading the Methods, I expected to see more of these details there and am a bit confused as to why they are included in the Results.
5. Suggest not paraphrasing Steve Jobs in the Discussion section.
6. The manuscript states that this pilot was conducted in 2018. The study also notes that ongoing work with 250 mother-infant pairs is currently being conducted, now 5 years later. Given how much has happened in the world, I am curious if the authors have any reflections on how the pandemic has changed the way we should understand and reflect their findings. (The pandemic need not be addressed in the manuscript, but the second to last paragraph of the Discussion talks about health emergencies. I am skeptical how such an involved pilot could be so quickly mobilized to respond to health emergencies. The authors should reflect on this if they believe findings point to this as a possibility. I also think the detailed statistics about flooding in Pakistan and other emergencies are out of scope for the paper. I don't believe anyone needs convincing that health emergencies of this nature exist.)

Round 2 Review

The authors have very thoughtfully and substantially revised the paper, making clear the methods and contributions of the study. I appreciate the detail with which the authors pointed to their edits in the revised manuscript and am satisfied with their changes.

At this point, I suggest only very minor revisions, asking authors to check grammar and conduct a copyedit of the paper. There are instances where a careful copyedit will improve the overall reading experience of the paper. For instance, in the Abstract, I suggest the following changes:

1. Drop the "the" in "Can the information-technology (IT) help these CHWs?"
2. Add a comma after "application" in "We explored answers through development and feasibility testing of Sehat Ghar, an android-based digital application to improve the communication capacity of volunteer CHWs in two slums of Islamabad."

3. Do not capitalize “Focus Group Discussions” in the Methods section.

Conflicts of Interest

None declared.

References

1. Haq ZU, Naeem A, Zaeem D, Sohail M, Pervaiz NUA. Development of a digital platform to promote mother and child health in underserved areas of a lower-middle-income country: mixed methods formative study. JMIRx Med 2024;5:e48213. [doi: 10.2196/48213]
2. Government of Pakistan. National Report of Pakistan for Habitat III. Habitat III. 2015 Jun. URL: https://habitat3.org/wp-content/uploads/Pakistan-Final-in-English.pdf [accessed 2024-06-24]

Abbreviations

CHW: community health worker
mHealth: mobile health

Edited by E Meinert; submitted 10.05.24; this is a non–peer-reviewed article;accepted 10.05.24; published 31.07.24.

Please cite as:
Johnston JS
Peer Review of “Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study”
JMIRx Med 2024;5:e60430
URL: https://xmed.jmir.org/2024/1/e60430
doi:10.2196/60430

Peer Review of “Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/49969

Companion article: https://www.medrxiv.org/content/10.1101/2023.06.01.23290351v1

Companion article: https://med.jmirx.org/2024/1/e55510

Companion article: https://med.jmirx.org/2024/1/e49969

(JMIRx Med 2024;5:e55727) doi:10.2196/55727

KEYWORDS

audiology; hearing; high-frequency; wristband; develop; development; wearable; wearables; machine learning; phoneme; phonemes; hear; vibrotactile; vibration; vibrations; sound; sounds; hearing loss; loud noise; loud noises; noise pollution; hearing aids; hearing aid

This is the peer-review report for “Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation”.

Round 1 Review

General Comments

This paper [1] highlights the utility and perceived communication benefits of the Clarity vibrotactile band for users with high-frequency hearing loss. Overall, this is a well-designed study that demonstrates the effectiveness of this assistive listening device that provides benefits for listeners with high-frequency hearing loss in complex listening situations as measured by the Abbreviated Profile of Hearing Aid Benefit (APHAB). Additionally, this study provides subjective evidence that both hearing aid (HA) users and non-HA users experience benefit from the Clarity device. Specifically, the non-HA users report more benefits across different listening conditions (background noise and reverberation) than HA users.

Specific Comments

Major Comments

1. Consider referencing Glick and Sharma [2] in your Introduction as it relates to the cross-modal plasticity associated with age-related hearing loss (presbycusis).
2. In the Methods section, consider starting with a clear description of the participants. Who are they, how many, how many were HA users versus non-HA users, age, etc. While the majority of this information is embedded later in the article, it is not readily accessible.
3. In the Methods section, consider creating a subheading or table for the audiometric data of the participants and including additional information like a description of their audiometric data (type, degree, configuration), pure tone average (500, 1000, and 2000 Hz), symmetry of the hearing loss, how many were considered to be within normal limits up to 2000 Hz versus having hearing loss at lower frequencies (≤2000 Hz). This could have a significant impact on speech understanding difficulties, especially in complex listening environments.
4. For the audiometric data, how many participants provided their test results from a doctor of audiology or hearing health care professional? How many provided results from the mobile app? Is it possible to confirm that all participants had sensorineural hearing loss and not mixed or conductive hearing loss?
5. In the Device subsection, consider adding additional information regarding the microphone characteristics. Additionally, define “GRMS.”
6. In the Algorithm subsection, you mention the sham algorithm and the /f/ motor. In the sham condition, which motor represents the /f/ phoneme, and which additional phonemes are used in the sham condition?
7. Additionally, the sham condition is never mentioned in the Results or Discussion. Consider adding this information to the manuscript, or if you choose not to, consider not introducing the sham algorithm.
8. In Figure 3, consider changing the y-axis to “APHAB Score (%)” and refer to the APHAB benefit scores as scores or percentages instead of points in the text.
9. For the simple linear regression, consider adding a statement that indicates what this means or its importance.
10. In Figure 5, consider adding bars for weeks 0 and 1 to help readers visualize the results in the text.

11. Consider creating a line graph that highlights the greater decrease in APHAB scores from baseline to week 6 for those without HAs than those with HAs (as discussed in the Results).
12. In Figure 6, this figure represents benefit scores from baseline (wk 0) to week 6, correct? Consider clarifying the figure text and removing the information regarding the subgroups.
13. In the Discussion and Conclusion sections, I do not think it is accurate to say that the Clarity device "improved their understanding of speech communication" because that was not what was measured. The APHAB is a subjective measure, which to me means that all the benefits users received from using the Clarity are perceived benefits and are not measurable improvements in understanding. To claim speech understanding improvements, I feel you would need to document that through an objective speech understanding measure such as the word recognition score in quiet, word recognition score in noise, Quick Speech in Noise, etc.
14. In the Discussion section, you refer to the group with a higher APHAB score experiencing a greater improvement. Is this the group that uses HAs, or is this a different subgroup? It would be interesting to know how many in this group had hearing loss between 250-2000 Hz.
15. In the Discussion section, you report subgroup data for background noise, reverberation, and ease of communication that is not documented or reported in the Results section or any figures/tables. Consider adding this.
16. In the Conclusion section, you mention that "results also demonstrate that individuals who had the greatest amount of difficulty understanding speech prior to." Is this the without HA subgroup or a different subgroup? A few times throughout the article, these labels appear to be used interchangeably. While this may be accurate for your data set, I would caution that these terms/labels are not mutually exclusive.

Minor Comments

1. In the Introduction, the authors mention that HA and cochlear implant users commonly report disappointment with understanding speech and reference Hickson et al [3]. While this could be true, the majority of users' complaints are specifically related to difficulties understanding speech in complex or noisy listening environments, not just in quiet as is implied.
2. How much were participants compensated for their participation?
3. In Figure 2, I assume your scale for the y-axis is dB of HL? Consider clarifying which dB scale was used.
4. In the Paradigms subsection, does the Clarity device have any data logging features that can objectively record how often or how long the participant is using the device or in what listening conditions the user is in with the device (eg, quiet rooms, noisy restaurants, or reverberant auditoriums)?
5. In the APHAB subsection, consider rewording for clarity: "modified version of the Abbreviated Profile of Hearing Aid Benefit (APHAB) which did not include six questions related to the aversiveness subscale (Cox, 1997)."
6. In the Results section, consider rewording for clarity: "...they ended the study at a lower level of disability than those with hearing aids."
7. The implication of microphone location briefly mentioned in the Discussion is very important in my opinion. Microphone location is a significant issue even for ear-level HAs. I can only imagine the microphone placement significantly impacts the benefit and utility of the Clarity.
8. In the Conclusion section, consider rewording for clarity: "We found that while both hearing aid and non-hearing aid users with high frequency hearing loss reported benefited, vibrotactile feedback appears to be more beneficial for non-hearing aid users."
9. The manuscript does not include an ethical approval statement or a limitations section.

Conflicts of Interest

None declared.

References

1. Kohler I, Perrotta MV, Ferreira T, Eagleman DM. Cross-modal sensory boosting to improve high-frequency hearing loss: device development and validation. JMIRx Med 2024;5:e49969. [doi: 10.2196/49969]
2. Glick H, Sharma A. Cross-modal plasticity in developmental and age-related hearing loss: clinical implications. Hear Res 2017 Jan;343:191-201. [doi: 10.1016/j.heares.2016.08.012] [Medline: 27613397]
3. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. Int J Audiol 2014 Feb;53 Suppl 1:S18-S27. [doi: 10.3109/14992027.2013.860488] [Medline: 24447233]

Abbreviations

APHAB: Abbreviated Profile of Hearing Aid Benefit
HA: hearing aid

Edited by E Meinert; submitted 21.12.23; this is a non–peer-reviewed article;accepted 21.12.23; published 09.02.24.

Please cite as:
Anonymous
Peer Review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"
JMIRx Med 2024;5:e55727
URL: https://xmed.jmir.org/2024/1/e55727
doi:10.2196/55727

Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45688

Companion article: https://med.jmirx.org/2024/1/e55863

Companion article: https://med.jmirx.org/2024/1/e45688

(JMIRx Med 2024;5:e56040) doi:10.2196/56040

KEYWORDS

COVID-19; injury; prevalence; adaptation; sports medicine; follow-up; training; football; epidemiology; sport; athlete; athletic; injuries

This is the peer-review report for "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study."

Round 1 Review

General Comments

This paper [1] provides epidemiological data on injury incidences in the National Football League (NFL) before and after the COVID-19 lockdown. This paper has the potential to be clinically meaningful; however, it has major flaws that should be addressed before it is reconsidered.

Specific Comments

Major Comments

1. The authors stated that "This is the first large scale opportunity to demonstrate the effects of these principles and how they are important to understanding injury epidemiology." However, there are studies that have looked at the effect of COVID-19 on other sporting leagues, for example, Bundesliga, Major League Soccer (MLS), etc. All this has been published and should be cited.
2. Can the authors please confirm or comment on the potential accuracy of this open data? What validity checks were employed to demonstrate that these data are accurate?
3. How can the authors be sure that the increase in injuries was due to COVID-19?
4. What is meant by injuries? Soft tissue injuries or concussions? Perhaps an analysis on time-loss injuries would be more beneficial and add value.

Conflicts of Interest

None declared.

Reference

1. Puga TB, Schafer J, Thiel G, et al. COVID-19 National Football League (NFL) injury analysis: follow-up study. JMIRx Med 2024;5:e45688. [doi: 10.2196/45688]

Abbreviations

MLS: Major League Soccer
NFL: National Football League

Edited by E Meinert; submitted 03.01.24; this is a non–peer-reviewed article;accepted 03.01.24; published 13.02.24.

Please cite as:
Anonymous
Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"
JMIRx Med 2024;5:e56040
URL: https://xmed.jmir.org/2024/1/e56040
doi:10.2196/56040

Peer Review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers"

Colleen Peterson, PhD

Transportation Research Institute (UMTRI), University of Michigan, Ann Arbor, MI, United States

Related Articles:

Companion article: http://preprints.jmir.org/preprint/42211

Companion article: https://www.medrxiv.org/content/10.1101/2022.05.25.22275517v2

Companion article: https://med.jmirx.org/2024/1/e56439

Companion article: https://med.jmirx.org/2024/1/e42211

(JMIRx Med 2024;5:e56444) doi:10.2196/56444

KEYWORDS

Latino; dementia; caregiving; COVID-19; Alzheimer's disease; health disparity; qualitative research; transcript analysis; health inequality; minority population; epidemiology; peer support; social services; health care; Alzheimer's; minority; qualitative; interview; caregiver; primary care; impact; resilience; disparity; outcome; Alzheimer disease; Alzheimer

This is the peer-review report for "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers."

Round 1 Review

General Comments

The authors [1] present a compelling argument for understanding how a doubly vulnerable population in the United States (Latino persons with dementia) experienced the COVID-19 pandemic. To this end, they share the results of a thematic analysis of interviews with primary care providers and caregivers of Latino persons living with dementia. The qualitative analysis could use better explanations and the themes could be more descriptive. Moreover, the many of themes do not capture or explain their relevance to understanding the intersectionality of dementia and Latino lives. My comments below speak to that, as well as other issues. The manuscript has a good foundation of an informative article that showcases the lived experience of this population during a critical time and could be modified to provide formative evidence for improving care, inside and outside of a pandemic.

Specific Comments

Major Comments

1. Consider making subthemes or codes more descriptive and meaningful. Good themes tell you what the story is or what the direction is at least (good or bad). Some of these are readily available (or easily modified) from sentences in the paper already (eg, "the pandemic influence[d] mental and emotional health"; "Social support was critical for reducing social isolation and its sequalae"; caregivers and persons with dementia "lost access to engaging activities during the confinement"; and "Remote communication facilitated social support"). See other qualitative research for examples. There are even good examples in the dementia care literature during the pandemic. For example:

Mitchell LL, Horn B, Stabler H, et al. Caring for a relative with dementia in long-term care during the COVID-19 pandemic: a prospective longitudinal study. Innov Aging. 2023 Apr 17;7(4):igad034. doi: 10.1093/geroni/igad034. PMID: 37213326; PMCID: PMC10195573.

Harding E, Rossi-Harries S, Gerritzen EV, et al. "I felt like I had been put on the shelf and forgotten about" – lasting lessons about the impact of COVID-19 on people affected by rarer dementias. BMC Geriatr. 2023;23(392). https://doi.org/10.1186/S12877-023-03992-1

2. Throughout: The topic is about the Latino Alzheimer disease and related dementias (ADRD) experience, but many themes do not tap into this overlap of the Latino and ADRD experience. It is not made relevant to the ADRD experience or it is not explained how the ADRD context affected it, either in the theme itself (eg, see "poor nutrition" codes) or in the quote used to justify it (eg, see "stress" and "work" codes). A thorough review of the codes and quotes to meet this intersectionality would be beneficial.

3. Facilitators and barriers are not often used as codes on their own but a way to categorize or further delve into aspects of other issues. What was facilitated? What was the barrier blocking? I could see weaving facilitators, barriers, and consequences into the discussions around the other codes, like informal and formal support.

4. Please clarify the methods.

a. I have not heard of using condensed transcripts before. Why was this done? What was taken out exactly? How were "meaningful bits of text" identified?
b. It is not clear who was doing the coding at which times. There is the first author as a coder and then 2 additional coders, but the final sentence indicates there were just 2.
c. Did the first author make all the themes and do all the coding and then the other coder(s) just reviewed it? (Rather than all independently reviewing and coming together to come up with themes and then independently coding and later addressing the coding discrepancies.)

5. Add headers in the text for each subtheme and the codes for better flow and to help readers keep track of what theme or code they are reading about.

6. For quotes in general:

a. Consider editing any quote over 2 lines or selecting briefer quotes. Three lines is okay if compelling. If more than 3, it has to be a really good quote. There are 2 very long quotes in "other impacts."
b. Provide some context or active linking to the code as lead-in text.
c. Make sure they are absolutely relevant to the Latino and ADRD experience.

7. For the discussion:

a. Consider adding comparisons of findings to non-Latino ADRD COVID-19 experiences to highlight the differences this population experienced and to better showcase why continued attention to this specific population is warranted. See the Mitchell et al and Harding et al papers cited above as potential comparison points.
b. Why is circular migration being brought up? As written, this does not appear to be ADRD related and was only lightly discussed in the results (though, it was not clear if it is ADRD related in the results either).
c. "The fact that some PCPs suspected an unknown longer-term impact of the COVID-19 pandemic warrants further longitudinal research into this topic." While true, it is strange to frame ongoing need to understand this based on participant responses alone. Also, it needs to be related to ADRD.
d. "Caregivers' reports on health care providers' confusion between ADRD and COVID-19 infection symptoms warrants research to improve diagnosis and severity assessments of both conditions." Where did this come from? It feels unsupported from the evidence provided in the current study and an odd place to end the discussion.

8. For the conclusion:

a. "This pandemic has revealed many of the barriers that Latino families with ADRD face, and in most cases, this has exacerbated previous barriers. However, with every crisis comes an opportunity for improvement, which will hopefully translate into improved conditions among Latino families with ADRD." This does not really say anything; be specific regarding the barriers and what could be improved. You could succinctly use the start of the next sentence to end this one.
b. "These improved conditions might include more equitable access to health care and community services, a better quality of these services, subsidized formal and informal supports, and flexible hybrid means of communication." Which would lead to…or mean what for the (public) health of Latino persons living with ADRD and their carers? What is the overall takeaway pertaining to health or public health?
c. The discussion and conclusions could be broadened out to medical care in general if the issues appear to also be independent of the pandemic.

Minor Comments

9. Mention the United States as the target population in the abstract.

10. Typos: "The fist author also condensed" and "work in the meat packing plan industry."

11. Clarify "To make bring rigor and validity." Or is there a typo here?

12. "To make bring rigor and validity to the research process, the interviewer used active listening techniques during the interview aimed at confirming the information shared by participants. The interviewer also emphasized the fact that participants were the experts in their experiences to reduce power differentials." This belongs in the methods, not analysis.

13. "[Explains how after the lockdown, the care recipient only remembers long term memories]. So, I was thinking that all the time she was locked down here because of the cold weather and COVID might have affected her more." Avoid total paraphrasing and provide (translated) direct quotes. Or put the paraphrase as context before the quote.

14. "...PCPs had to reduce physical contact with care recipients, which reduced their chance to convey warmth to their patients." A quote here would be nice.

15. "This fear was not unfounded. Prior to the availability of the vaccine, caregivers and care recipients acquired COVID-19. As Latino older adults, they were at an increased risk for complications including death, causing significant chronic concern and fear." This is useful for the introduction (and the vulnerability was discussed) and discussion but should not be a part of the results. I suggest omitting this.

16. "Fourth, care recipients and PCPs highlighted the frequency and severity of depressed mood among caregivers and care recipients, especially during the lockdown due to lack of social support and social isolation. The PCPs noted that lack of social support and social isolation due to lockdown negatively impacted mood, sharing:" This is redundant. Consider condensing into 1 sentence.

17. “Fourth, consequences of social support were physical, psychological, and social. Examples of physical consequences include potentially reducing mortality by providing formal and informal caregiving services. Psychological consequences include clinic and family support reducing loneliness and increasing feelings of safety. Social consequences include caregivers being allowed to accompany their care recipients during clinic visits, curbside visits allowing socialization and home care services lowering isolation.” Like the barriers above this section, these feel more like they could be part of the informal or formal support codes. What are the supporting quotes? Also, it is not clear what were the consequences? What were the causes that led to the consequences?

18. “Third, consequences of the higher use of remote communication were both positive and negative.” This should just be part of the remote communication theme description.

19. “...similar to other studies, the need to rely on remote communication intensified the digital divide.” Reverse it—the digital divide was problematic given the need to rely on remote communication.

20. “...for their survival” in the conclusion is a bit heavy-handed. Speak on something closer at hand in the manuscript like avoiding exposure and infection.

Round 2 Review

General Comments

The authors thoroughly attended to the reviewer responses. The methods are easy to understand and the rework of the themes and related quotes in the results is a great improvement. The discussion could be easier to read by breaking the large paragraphs into smaller ones. The conclusion should be revised to more specifically attend to what the study found and how it extends the literature. These and other issues are further noted:

Specific Comments

Major Comments

Major comments in order of appearance in the manuscript:

1. In the 2.2. poor nutrition theme, it is more obvious that the mailing system and financial insecurity could directly be influenced by the pandemic, but it is not clear that skills and level of impairment were affected by the pandemic. As written, it sounds more like an overarching ADRD problem rather than an ADRD issue specific to the pandemic. This should be clarified, especially since it is brought up specifically in the discussion as a unique finding.
2. Consider reworking this sentence to clarify and streamline: “While home-delivered meals operated normally, Latino families with ADRD tried to access these for the first time during the pandemic to obtain food while reducing the risk of infection.” My suggested rework is “Some Latino families with ADRD we interviewed tried to use home-delivered meals for the first time during the pandemic to reduce risk of infection.”
3. Break large paragraphs throughout the discussion into smaller ones by more specific topic (eg, food and nutrition, work changes, and infection risk).
4. How exactly are fatalism and personalism related to the findings in the study? Make an explicit tie back into the findings to make a stronger ending to this part of the discussion.
5. Rephrase “Was this also the case for cognitive and functional decline?” into a statement rather than a question.
6. As written, the conclusion paragraph does not indicate well what the study found or how it extends the literature. The first sentence needs to be reworked—who is “their” referring to? Families were critical to “maintaining or improving “health and quality of life,” correct? Make succinct and specific mention to how the families were affected “beyond infection and physical symptoms.” What were the specific barriers that were exacerbated?

Minor Comments

1. “Other caregivers or their care recipient had been infected or were indeed infected during the interview.” This sounds like the interviewer infected them. They were experiencing COVID-19 at the time of the interview?
2. Typo in theme 4.3: “to the their.”
3. Avoid the numeric two in the quote in theme 5.3: “Mom had 2 that got COVID.” Suggested rework: “Mom had two [home assistants] that got COVID.”
4. Remove the hyphen from “frequently-mentioned.”

Conflicts of Interest

None declared.

Reference

1. Perales-Puchalt J, Peltzer J, Fracachan-Cabrera M, et al. Impact of the COVID-19 pandemic on Latino families with Alzheimer disease and related dementias: qualitative interviews with family caregivers and primary care providers. JMIRx Med 2024;5:e42211. [doi: 10.2196/42211]

Abbreviations

ADRD: Alzheimer disease and related dementias

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 08.03.24.

Please cite as:
Peterson C
Peer Review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers"
JMIRx Med 2024;5:e56444
URL: https://xmed.jmir.org/2024/1/e56444
doi:10.2196/56444

Peer Review of "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020"

Anonymous

Related Articles:

Companion article: http://preprints.jmir.org/preprint/44381

Companion article: https://med.jmirx.org/2024/1/e57620

Companion article: https://med.jmirx.org/2024/1/e44381

(JMIRx Med 2024;5:e57935) doi:10.2196/57935

KEYWORDS

access to health; health service utilization; eye care use; health-seeking behavior; sociodemographic determinant; visual impairment; social support; women empowerment; education; eye care; pediatric; eye; ophthalmology; visual; ECU; eye service; utilization; Malawi; empowerment; health service use; use

This is the peer-review report for "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020."

Round 1 Review

General Comments

This paper [1] is a secondary analysis of a Malawian household survey exploring associations of patients who self-reported as having used formal eye care services. It is a useful idea to use this survey data for this purpose, but the author needs to check that they are using the correct source numbers for their statistical analysis and only report the numbers actually surveyed—not the national estimated numbers derived from these.

Specific Comments

Major Comments

1. "In Malawi, 3.3% of the population is blind compared to 1.01% in America [2,3]." There is no way 3% of Malawi is blind. (Half of Malawi's population are children, so if 3% of Malawi was blind, that would be about 1 in 20 adults—not possible.) Check your references.

2. The abstract needs improving to give the definition of eye care use (ECU). In the results, it says "The prevalence of ECU was 60.6%," which is not really a prevalence unless you give a clearer definition, for example "of those with eye symptoms, what proportion have access formal eye care services in the two weeks prior to the survey date."

3. The sample was 28,388 adults? You cannot, then, in the results' "Characteristics of study participants" section say there were 6 million young adults involved or that 5,660,836 (56%) of the adults were married. You also can't say that "27,336 (0.3% of 2,734,768) complained of ocular symptoms." This is the main problem with the report—you need to give the actual numbers of people surveyed who reported ocular symptoms—presumably 0.3% of 28,388—which is only 85 people. Thus your CIs/other statistical analyses around estimates with a sample of 85 people reporting eye symptoms will be quite different than if you extrapolate to the whole population of Malawi.

Minor Comments

4. "We entered the variables...": Who is "we"? I only see one author

5. "Sort care" should be "sought care": This is used 5 times in the paper so should be changed at all uses

6. There are some random capital letters in various places: "that In Malawi"—why has "In" got a capital?

Conflicts of Interest

None declared.

References

1. Mzumara T, Kantaris M, Afonne J. Eye care service use and associated health-seeking behaviors among Malawian adults: secondary analysis of the Malawi Fifth Integrated Household Survey 2019-2020. JMIRx Med 2024;5:e44381. [doi: 10.2196/44381]
2. Kalua K, Lindfield R, Mtupanyama M, Mtumodzi D, Msiska V. Findings from a rapid assessment of avoidable blindness (RAAB) in Southern Malawi. PLoS One 2011 Apr 25;6(4):e19226. [doi: 10.1371/journal.pone.0019226] [Medline: 21547074]
3. Flaxman AD, Wittenborn JS, Robalik T, et al. Prevalence of visual acuity loss or blindness in the US: a Bayesian meta-analysis. JAMA Ophthalmol 2021 Jul 1;139(7):717-723. [doi: 10.1001/jamaophthalmol.2021.0527] [Medline: 33983373]

Abbreviations

ECU: eye care use

Edited by E Meinert; submitted 29.02.24; this is a non–peer-reviewed article;accepted 29.02.24; published 09.04.24.

Please cite as:
Anonymous
Peer Review of "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020"
JMIRx Med 2024;5:e57935
URL: https://xmed.jmir.org/2024/1/e57935
doi:10.2196/57935

Peer Review of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study"

Mario Coccia

National Research Council of Italy, Turin, Italy

Related Articles:

Companion article: http://preprints.jmir.org/preprint/43341

Companion article: https://med.jmirx.org/2024/1/e59637

Companion article: https://med.jmirx.org/2024/1/e43341

(JMIRx Med 2024;5:e59638) doi:10.2196/59638

KEYWORDS

COVID-19; pandemic; health care disparities; intensive care unit; ICU; right to health; quality of care; intubation; mortality; health disparity; health inequality; surveillance data; in-patient; mortality; COVID-19 patient; hospitalization; disparity; inequality; surveillance; health care system; Greece; region; Delta; Omicron; vaccination; vaccine; public health; patient load; deterioration; time

This is the peer-review report for "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study."

Round 1 Review

General and Detailed Comments

The topics of this paper [1] are interesting, though well known.

The structure and content must be revised, and results have to be better explained by authors before being reconsidered for publication.

Title has to be shorter.

Abstract has to clarify the goal, sample, results, and health and social implications to cope with the next pandemics to improve health care.

Introduction is poor and has to better clarify the research questions of this study and provide more theoretical background about strategies of prevention and good governance to cope with the pandemic crisis. After that they can focus on the topics of this study to provide a correct analysis for fruitful discussion (see suggested readings that must be all read and used in the text).

Methods of this study are not clear. The section of Materials and Methods must be restructured with the following 3 points in the same order:

- Sample and data
- Measures of variables
- Models and data analysis procedure

Results: It is not clear why authors apply a Mann-Whitney test, considering the large sample it is better to apply an independent sample *t* test, or some other parametric test. This additional test can support better results if they are reliable. Table 1: avoid acronyms; specify ICU as intensive care unit. Same comments for Figure 2.

Discussion: first, authors have to synthesize the main results in a simple table to be clear for readers and then show what this study adds compared to other studies. In addition, authors should discuss the type of mechanical ventilation, because studies show that invasive ventilation (intubation) creates VAP (ventilator-associated pneumonia), and a lot of people intubated for COVID-19 died from this problem rather than COVID-19. Countries that have reduced mortality, such as Germany and New Zealand, used mainly noninvasive ventilation, which can better treat patients and avoid mortality with new technology. See suggested papers.

Conclusion has to be added as an autonomous section. Conclusion not to be a summary, but authors have to focus on manifold limitations of this study and provide suggestions for health, crisis management, and social policy, as well as how nations can prevent, with good governance and new technology in artificial ventilation, the next pandemics and improve health care with vaccination, noninvasive ventilation, and nonpharmaceutical measures of control. In this manner the paper can provide useful policy implications for Greece and improve health care for the next pandemics.

Overall, then, the paper is interesting. The theoretical framework is weak, and some results create confusion...the structure of the paper has to be improved; study design, discussion, and presentation of results have to be clarified using the suggested comments.

I strongly suggest improving the paper by using all comments (suggested papers that are included to all be read and used) that I will verify in-depth, and maybe it can be considered. If the paper is not improved as suggested it will be dismissed.

Suggested readings of relevant papers that have to be read and all inserted in the text and references:

Nasrullah A, Gangu K, Garg I, et al. Trends in hospitalization and mortality for influenza and other respiratory viruses during the COVID-19 pandemic in the United States. Vaccines (Basel). Feb 10, 2023;11(2):412.

Coccia M. High potential of technology to face new respiratory viruses: mechanical ventilation devices for effective healthcare to next pandemic emergencies. Technol Soc. May 2023;73:102233.

COVID-19 Forecasting Team. Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. Lancet. Apr 16, 2022;399(10334):1469-1488.

Benati I, Coccia M. Effective contact tracing system minimizes COVID-19 related infections and deaths: policy lessons to reduce the impact of future pandemic diseases. J Pub Admin Gov. Sept 2022;12(3):19-33.

Velissaris D, Paraskevas T, Oikonomou E, Bizos A, Karamouzos V, Marangos M. Evaluation of four novel prognostic scores on admission for COVID-19 mortality. An experience from a Mediterranean tertiary center. Acta Clin Belg. Aug 2022;77(4):748-752.

Magazzino C, Mele M, Coccia M. A machine learning algorithm to analyse the effects of vaccination on COVID-19 mortality. Epidemiol Infect. Sept 12, 2022;150:e168.

Kondilis E, Tarantilis F, Benos A. Essential public healthcare services utilization and excess non-COVID-19 mortality in Greece. Public Health. Sep 2021;198:85-88.

Coccia M. Sources, diffusion and prediction in COVID-19 pandemic: lessons learned to face next health emergency. AIMS Public Health. Mar 2, 2023;10(1):145-168.

Mazzucchelli R, Agudo Dieguez A, Dieguez Costa EM, Crespí Villarías N. Democracia y mortalidad por COVID-19 en Europa. Rev Esp Salud Publica. Jun 24, 2020;94:e202006073.

Coccia M. Preparedness of countries to face COVID-19 pandemic crisis: strategic positioning and underlying structural factors to support strategies of prevention of pandemic threats. Environ Res. Jan 2022;203:111678.

Marinaki C, Kapadochos T, Katsoulas T, et al. Estimation of the optimal time needed for weaning of intensive care unit tracheostomized patients on mechanical ventilation. A prospective observational study. Acta Biomed. Apr 24, 2023;94(2):e2023103.

Coccia M. Optimal levels of vaccination to reduce COVID-19 infected individuals and deaths: a global analysis. Environ Res. Mar 2022;204(Pt C):112314.

Maltezou HC, Basoulis D, Bonelis K, et al. Effectiveness of full (booster) COVID-19 vaccination against severe outcomes and work absenteeism in hospitalized patients with COVID-19 during the Delta and Omicron waves in Greece. Vaccine. Mar 31, 2023;41(14):2343-2348.

Benati I, Coccia M. Global analysis of timely COVID-19 vaccinations: improving governance to reinforce response policies for pandemic crises. Int J Health Governance. Aug 12, 2022;27(3):240-253.

Stasinos N, Kousis A, Sarlis V, et al. A tri-model prediction approach for COVID-19 ICU bed occupancy: a case study. Algorithms. 2023;16(3):140.

Coccia M. Comparative critical decisions in management. In: Farazmand A, editor. Global Encyclopedia of Public Administration, Public Policy, and Governance. Springer Nature. 2020.

Aslanidis V, Tsolaki V, Papadonta ME, et al. The impact of the COVID-19 pandemic on mental health and quality of life in COVID-19 department healthcare workers in Central Greece. J Pers Med. Jan 29, 2023;13(2):250.

Coccia M. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Sci Total Environ. Aug 10, 2020;729:138474.

Panagiotopoulos ΦΙ. Corporate social responsibility initiatives and programs in the health system of Greece due to the pandemic of COVID-19. In: Indowu MT, Idowu AO, editors. Corporate Social Responsibility in the Health Sector. CSR, Sustainability, Ethics & Governance. Springer. 2023:93-110.

Coccia M. COVID-19 pandemic over 2020 (with lock downs) and 2021 (with vaccinations): similar effects for seasonality and environmental factors. Environ Res. May 15, 2022;208:112711.

Malli F, Lampropoulos IC, Perlepe G, Papagiannis D, Gourgoulianis KI. Analysis of SARS-CoV-2 cases, COVID-19 outcomes and vaccinations, during the different SARS-CoV-2 variants in Greece. Vaccines (Basel). Jan 4, 2023;11(1):126.

Coccia M. Pandemic prevention: lessons from COVID-19. Encyclopedia. 2021;1(2):433-444.

Gavrielatou E, Vaporidi K, Tsolaki V, et al. Rapidly improving acute respiratory distress syndrome in COVID-19: a multi-centre observational study. Respir Res. Apr 14, 2022;23(1):94.

Coccia M. Effects of strict containment policies on COVID-19 pandemic crisis: lessons to cope with next pandemic impacts. Environ Sci Pollut Res Int. Jan 2023;30(1):2020-2028.

Stafylaki D, Maraki S, Vaporidi K, et al. Impact of molecular syndromic diagnosis of severe pneumonia in the management of critically ill patients. Microbiol Spectr. Oct 26, 2022;10(5):e0161622.

Coccia M. Covid-19 vaccination is not a sufficient public policy to face crisis management of next pandemic threats. Public Organiz Rev. Oct 17, 2022;23:1353-1367.

Rapti I, Asimakopoulos A, Liontos A, et al. Association of patient characteristics with clinical outcomes in a cohort of hospitalised patients with SARS-CoV-2 infection in a Greek referral centre for COVID-19. Epidemiol Infect. Aug 16, 2022;150:e160.

Coccia M. Improving preparedness for next pandemics: max level of COVID-19 vaccinations without social impositions to design effective health policy and avoid flawed democracies. Environ Res. Oct 2022;213:113566.

Lytras T, Tsiodras S. Total patient load, regional disparities and in-hospital mortality of intubated COVID-19 patients in Greece, from September 2020 to May 2021. Scand J Public Health. Aug 2022;50(6):671-675.

Editorial Notice

This peer-review report was modified upon sending to the authors due to editorial policies.

Conflicts of Interest

None declared.

Reference

1. Lytras T. Health care system overstretch and in-hospital mortality of intubated patients with COVID-19 in Greece from September 2020 to April 2022: updated retrospective cohort study. JMIRx Med 2024;5:e43341. [doi: 10.2196/43341]

Edited by E Meinert; submitted 17.04.24; this is a non–peer-reviewed article;accepted 17.04.24; published 10.06.24.

Please cite as:
Coccia M
Peer Review of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study"
JMIRx Med 2024;5:e59638
URL: https://xmed.jmir.org/2024/1/e59638
doi:10.2196/59638

Peer Review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"

Robert Eikelboom[1,2,3,4], BEng, MSc, PhD

1

2

3

4

Related Articles:

Companion article: https://preprints.jmir.org/preprint/49969

Companion article: https://www.medrxiv.org/content/10.1101/2023.06.01.23290351v1

Companion article: https://med.jmirx.org/2024/1/e55510

Companion article: https://med.jmirx.org/2024/1/e49969

(JMIRx Med 2024;5:e55554) doi:10.2196/55554

KEYWORDS

audiology; hearing; high-frequency; wristband; develop; development; wearable; wearables; machine learning; phoneme; phonemes; hear; vibrotactile; vibration; vibrations; sound; sounds; hearing loss; loud noise; loud noises; noise pollution; hearing aids; hearing aid

This is the peer-review report for "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation".

Round 1 Review

The authors report on an interesting study [1] in which they use a wearable device to sense high-frequency sounds. I have some specific comments below. To summarize, some essential elements are missing from the manuscript, and the manuscript needs significant editorial attention (errors, academic writing style, figures).

Introduction: I would suggest using primary references for the number of people with hearing loss (rather than Olusanya et al [2]) and for the burden of hearing loss (rather than Michels et al [3]). Regarding the risk of high-frequency hearing loss, have the authors overlooked the fact that this is commonly seen in most older adults (ie, what is attributed to aging)? This is mentioned in the second paragraph. The authors are mixing up noise-related hearing loss and age-related hearing loss (presbycusis) in the manuscript.

I do not think that Hickson et al [4] is a primary reference for limitations of hearing aids (HAs) and cochlear implants.

"The auditory cortex is activated by vibrotactile information in individuals who are hearing impaired and deaf." This implies that the auditory cortex is only activated this way.

Middle paragraph: phonemes are extracted. How this is done should be provided here, not later in the manuscript.

Is the designation of the particular transducer important? In other words, is a larger temporal difference between the two most similar phonemes important?

"The user is then able to understand..." Isn't this yet to be shown, or is evidence provided in the next paragraph? If so, this needs to be made clearer.

Interestingly, the microphone is placed on the wrist, a part of the body that can often be situated away from the direct line of communication between two people (eg, under a table). Were the users trained to keep their wrists up?

"...listening to an audiobook, podcast..." These are often streamed to personal headsets/earphones. Were any instructions provided in terms of volume, closeness to speakers, etc?

Participants: Normally, information about participants is provided before most of the other information in a Methodology section, particularly before, for example, tasks.

Abbreviated Profile of Hearing Aid Benefit (APHAB): I am not sure that I agree with the rationale that questions on aversiveness are not relevant. Cox and Alexander [5] write "Aversiveness of Sounds, quantifies negative reactions to environmental sounds," and "The APHAB is a potentially valuable clinical instrument. It can be useful for quantifying the disability associated with a hearing loss and the reduction of

disability that is achieved with a hearing aid." That is, it is designed to be used before an intervention (and has been used a lot for non-HA interventions as well, eg, implants).

How was the APHAB administered?

How many male and female participants were in the study?

Using an audiogram from any mobile-based device means little guarantee of accuracy.

What was the rationale for the specifications for the audiogram?

Any reason why 16 people were recruited?

What is "(11.6)," SD? And "(13)" and "(9)"?

Figure 3: I suggest not including the values on the plot. Furthermore, "Error boundary represents standard error of the mean." The reader has to interpret the "error boundary" as the gray area.

"...to drop at a slower, more steady pace for the remaining five weeks of the study." Writing could be tightened up a bit, and there is a rise in scores at 3 weeks. If the response is that there is not a significant increase, then it would be good to report at what point the difference is not significant.

Regression analysis: This is OK, but the use of a paired sample *t* test could have been taken for both analyses.

On the other hand, a multinomial regression analysis could have considered the influence of age, HA user or not, or baseline APHAB scores on final APHAB scores.

I see that there were approximately equal numbers of HA and non-HA users. Was this by accident or design? It is not mentioned in the Recruitment section.

Figure 6: It is not clear which score is being reported. At 6 weeks? I suspect it means the difference between the baseline and final scores. If so, this needs to be made clear in the caption.

It appears that there was no attempt to record the listening environments of the users nor how often they used their devices.

"Participants without hearing aids benefitted the most from vibrotactile sensory substitution..." True—in fact, those with HAs did not get significant benefits.

It is always good to devote a bit of space to the limitations of the study. This is missing in this manuscript.

"Future studies will focus on quantifying the maximum benefits possible and how long improvements continue before a plateau is reached." This is not a conclusion of the study.

Perhaps this is mentioned elsewhere, but the device is given a name; it would be good to know about the association between the authors and the manufacturer of the device.

Round 2 Review

General Comments

The authors appear to have responded to previous comments. However, having two different versions of the manuscript in the system has caused confusion. Having some sort of system to track changes would also have been very useful.

Specific Comments

The authors have persisted with using Michels et al [3]; this is not a primary reference for results of noise exposure of burden of hearing loss.

I am not convinced that even an omnidirectional microphone would be optimally placed on the wrist.

"...allow them to enjoy audio based entertainment such as movies and podcasts..." was of course not tested.

The last paragraph of the Introduction reads like a conclusion, not the presentation of aims or objectives.

I am unconvinced about the rationale for removing aversiveness from the APHAB; the same can be said about the other subscales. It is not about the unpleasantness introduced by the device; otherwise, why should the APHAB be applied before an intervention such as HAs or cochlear implants (as done in this study)? It is the person's overall aversiveness to sound. Anyway, the data were not collected, so there is little to be done.

"What was the rationale for the specifications for the audiogram?"

"This was simply a general inclusion criterion to make certain we were capturing garden-variety presbycusis."

It would be useful for this to be mentioned.

"The authors are associated both with Stanford University and the company Neosensory, which makes this device. This information is in the paper." Okay, but I think this should extend to more than noting the affiliation of the authors. Is a financial disclosure required?

Conflicts of Interest

None declared.

References

1. Kohler I, Perrotta MV, Ferreira T, Eagleman DM. Cross-modal sensory boosting to improve high-frequency hearing loss: device development and validation. JMIRx Med 2024;5:e49969. [doi: 10.2196/49969]
2. Olusanya BO, Davis AC, Hoffman HJ. Hearing loss: rising prevalence and impact. Bull World Health Organ 2019 Oct 1;97(10):646-646A. [doi: 10.2471/BLT.19.224683] [Medline: 31656325]

3. Michels TC, Duffy MT, Rogers DJ. Hearing loss in adults: differential diagnosis and treatment. Am Fam Physician 2019 Jul 15;100(2):98-108. [Medline: 31305044]
4. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. Int J Audiol 2014 Feb;53 Suppl 1:S18-S27. [doi: 10.3109/14992027.2013.860488] [Medline: 24447233]
5. Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. Ear Hear 1995 Apr;16(2):176-186. [doi: 10.1097/00003446-199504000-00005] [Medline: 7789669]

Abbreviations

APHAB: Abbreviated Profile of Hearing Aid Benefit
HA: hearing aid

Edited by E Meinert; submitted 15.12.23; this is a non–peer-reviewed article;accepted 15.12.23; published 09.02.24.

Please cite as:
Eikelboom R
Peer Review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"
JMIRx Med 2024;5:e55554
URL: https://xmed.jmir.org/2024/1/e55554
doi:10.2196/55554

Peer Review of “Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers”

Anonymous

Related Articles:

Companion article: http://preprints.jmir.org/preprint/42211

Companion article: https://www.medrxiv.org/content/10.1101/2022.05.25.22275517v2

Companion article: https://med.jmirx.org/2024/1/e56439

Companion article: https://med.jmirx.org/2024/1/e42211

(JMIRx Med 2024;5:e57160) doi:10.2196/57160

KEYWORDS

Latino; dementia; caregiving; COVID-19; Alzheimer’s disease; health disparity; qualitative research; transcript analysis; health inequality; minority population; epidemiology; peer support; social services; health care; Alzheimer’s; minority; qualitative; interview; caregiver; primary care; impact; resilience; disparity; outcome; Alzheimer disease; Alzheimer

This is the peer-review report for “Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers.”

Round 1 Review

General Comments

This paper [1] describes a qualitative study in which the authors seek to understand the experiences of Latino families managing Alzheimer disease and related dementias (ADRD). The authors interviewed both family caregivers and primary care providers (PCPs). This is a well-written manuscript that focuses on caregiving during COVID-19, a relatively understudied area. The authors note that this study represents secondary analyses of a larger study that focused on improving ADRD care services in primary care across settings.

Specific Comments

Major Comments

1. My key methodological critique is that the authors should follow the COREQ (Consolidated Criteria for Reporting Qualitative Research) recommendations in reporting this study. Since this is a completed study, it is possible that the authors will not meet all the criteria; however, it is still important to know which recommendations were followed and which were not.
2. My key conceptual critique is that the authors do not provide any rationale as to why family caregivers’ and PCPs’ perspectives are included together. While I certainly understand that these are secondary analyses, the study introduction still needs to justify why these 2 stakeholder groups will provide perspectives that can be synthesized.
3. Relatedly, the authors should provide a conceptual or theoretical framework that guided this study.
4. In methods, please clarify whether the main interview was 45-60 minutes and the additional COVID-19–specific questions were excluded from this time. Please also provide the interview guide as an appendix (see COREQ) and include sample questions in the manuscript.
5. Please clarify what is meant by “the interviewers emphasized that participants were the experts to reduce power differentials.” How was this accomplished? How did the interviewer ensure that this power differential existed and that it was subsequently reduced?
6. Consider expanding the description of theme 1 to capture the dimensions of impact noted in the subthemes, for instance, “Both caregivers and PCPs highlighted the physical, psychological and social impacts of the pandemic on patients with ADRD.”
7. The second theme does not appear to integrate the PCP and caregiver views as well as theme 1. Consider splitting the current theme 2 into 2: theme 2 could be about individual coping and resilience, and theme 3 could be about systems-level factors, which would include vaccination acceptance and remote communications.
8. The conclusion discusses death and formal care, which are not highlighted in the themes or results. Anchor the discussion to the results of this study.

Minor Comments

1. Please review and correct typographical errors; for example, in data analysis, it says “fist author” instead of “first author.”
2. The citation #16 is for focus groups, but the authors did 1:1 interviews. Please ensure that this citation is correct.
3. The percentages in Table 1 are not meaningful given the small sample size. Please report only the Ns.

Conflicts of Interest

None declared.

Reference

1. Perales-Puchalt J, Peltzer J, Fracachan-Cabrera M, et al. Impact of the COVID-19 pandemic on Latino families with Alzheimer disease and related dementias: qualitative interviews with family caregivers and primary care providers. JMIRx Med 2024;5:e42211. [doi: 10.2196/42211]

Abbreviations

ADRD: Alzheimer disease and related dementias
COREQ: Consolidated Criteria for Reporting Qualitative Research
PCP: primary care provider

Edited by E Meinert; submitted 06.02.24; this is a non–peer-reviewed article;accepted 06.02.24; published 08.03.24.

Please cite as:
Anonymous
Peer Review of “Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers”
JMIRx Med 2024;5:e57160
URL: https://xmed.jmir.org/2024/1/e57160
doi:10.2196/57160

Peer Review of “Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020”

Anonymous

Related Articles:

Companion article: http://preprints.jmir.org/preprint/44381

Companion article: https://med.jmirx.org/2024/1/e57620

Companion article: https://med.jmirx.org/2024/1/e44381

(JMIRx Med 2024;5:e58361) doi:10.2196/58361

KEYWORDS

access to health; health service utilization; eye care use; health-seeking behavior; sociodemographic determinant; visual impairment; social support; women empowerment; education; eye care; pediatric; eye; ophthalmology; visual; ECU; eye service; utilization; Malawi; empowerment; health service use; use

This is the peer-review report for “Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020.”

Round 1 Review

Potentially an interesting paper [1], but more details are needed in the methods to enable the reader to understand the results.

The English is not good throughout the document, and the writing could be much more succinct and precise.

Conflicts of Interest

None declared.

Reference

1. Mzumara T, Kantaris M, Afonne J. Eye care service use and associated health-seeking behaviors among Malawian adults: secondary analysis of the Malawi Fifth Integrated Household Survey 2019-2020. JMIRx Med 2024;5:e44381. [doi: 10.2196/44381]

Edited by E Meinert; submitted 13.03.24; this is a non–peer-reviewed article;accepted 13.03.24; published 09.04.24.

Please cite as:
Anonymous
Peer Review of “Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020”
JMIRx Med 2024;5:e58361
URL: https://xmed.jmir.org/2024/1/e58361
doi:10.2196/58361

Peer Review for “Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/55976

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.19.23298750v1

Companion article: https://med.jmirx.org/2024/1/e58273

Companion article: https://med.jmirx.org/2024/1/e55976

(JMIRx Med 2024;5:e59120) doi:10.2196/59120

KEYWORDS

monotherapy; GLP; FDA; adverse; reporting; thyroid; cancer; oncology; cancers; neoplasm; neoplasms; drug; drugs; pharmacy; pharmacies; pharmacology; pharmacotherapy; pharmaceutic; pharmaceutics; pharmaceuticals; pharmaceutical; medication; medications; glucagon-like peptide-1; Food and Drug Administration

This is the peer-review report for “Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis.”

Round 1 Review

General Comments

In this manuscript titled “Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis” [1], the authors analyzed over 18 million reports from the Food and Drug Administration (FDA) Adverse Event Reporting System, among which over 17,000 cases were identified to have increased possibility of thyroid hyperplasia and neoplasm when taking glucagon-like peptide-1 (GLP-1) receptor agonist (RA) monotherapy. The data were compared to the cases where the patients were taking sodium-glucose cotransporter-2 (SGLT-2) inhibitor monotherapy. Please see my suggestions and concerns below.

Suggestions and Concerns

1. The authors compared data between GLP-1 RA monotherapy and SGLT-2 inhibitor monotherapy. It would be great to have a little bit more introduction or description for the readers to understand why they compared the data with SGLT-2 inhibitor monotherapy. What is the function of SGLT-2 inhibitors during the therapy? I realized that the authors described this a bit in the *Methods* section, but more information would be appreciated to be added in the *Introduction* section for the readers to understand the rationale.
2. In Table 1, how are the unique individual thyroid hyperplasia and/or thyroid neoplasm–related adverse event (AE) reports being counted? Was it the sum of the cases number searched by the above AE Preferred Term in each section, or it was counted through searching a specific AE Preferred Term?
3. In the reporting odds ratio (ROR) analysis, where ROR = (a/b) / (c/d), I assume a indicates the number of AE cases in the exposed group, b indicates the number of non-AE cases in the exposed group, c indicates the number of AE cases in the control group, and d indicates the number of non-AE cases in control group. Is this correct? For calculating the ROR for all GLP-1 RAs (n=17,653; number of AEs=191) compared to the SGLT-2 inhibitor (n=14,102; number of AEs=7), maybe I am wrong, but should the ROR = (191 / [17,653 − 191]) / (7 / [14,102 − 7]), where (17,653 − 191) and (14,102 − 7) are the non-AE cases, which equals to 22.02? Did the authors use all the cases instead of the non-AE cases for the calculations? The same applies to the other ROR numbers and 95% CIs.
4. The authors claimed that GLP-1 RA monotherapy reports manifested a statistically significant increase in thyroid hyperplasia and neoplasm AEs when compared to SGLT-2 inhibitors. How was the statistical significance determined? Was it because the calculated ROR is over 1 (or 10) or the interval (ROR [−], ROR [+]) is large?
5. Figure 1’s resolution seems low in the document.

Conflicts of Interest

None declared.

Reference

1. Makunts T, Joulfayan H, Abagyan R. Thyroid hyperplasia and neoplasm adverse events associated with glucagon-like peptide-1 receptor agonists in the Food and Drug Administration Adverse Event Reporting System: retrospective analysis. JMIRx Med 2024;5:e55976. [doi: 10.2196/55976]

Abbreviations

AE: adverse event
FDA: Food and Drug Administration
GLP-1: glucagon-like peptide-1
RA: receptor agonist
ROR: reporting odds ratio
SGLT-2: sodium-glucose cotransporter-2

Edited by E Meinert; submitted 02.04.24; this is a non–peer-reviewed article;accepted 02.04.24; published 01.05.24.

Please cite as:
Anonymous
Peer Review for "Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis"
JMIRx Med 2024;5:e59120
URL: https://xmed.jmir.org/2024/1/e59120
doi:10.2196/59120

Peer Review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/49969

Companion article: https://www.medrxiv.org/content/10.1101/2023.06.01.23290351v1

Companion article: https://med.jmirx.org/2024/1/e55510

Companion article: https://med.jmirx.org/2024/1/e49969

(JMIRx Med 2024;5:e55728) doi:10.2196/55728

KEYWORDS

audiology; hearing; high-frequency; wristband; develop; development; wearable; wearables; machine learning; phoneme; phonemes; hear; vibrotactile; vibration; vibrations; sound; sounds; hearing loss; loud noise; loud noises; noise pollution; hearing aids; hearing aid

This is the peer-review report for "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation".

Round 1 Review

Overall

- This study [1] reports on an interesting device with intriguing clinical implications for people with hearing loss.
- Innovative, and worthy of reporting on this technology, which could inspire other researchers
- But there are some issues that I feel require revisions:
  - Conflation of self-reported and objective benefit in the write-up
  - Lack of reporting the range and dispersion of the data—paper focuses on group means and gives very little ability to draw any inferences about individual participant variability
  - Lack of objective data about performance of the algorithm and participant performance for speech understanding
  - No data presented for the final questionnaire presented in the Methods
  - Presentation and discussion of results switches back and forth between benefit scores and raw scores in a way that is unclear and makes the paper difficult to follow and interpret at times
  - Some conclusions are presented without statistical results to support them
  - Some conclusions are stated too strongly given the sample size and study design
  - Lack of a limitations section to help reader contextualize the results

Abstract

- "...improve their understanding of verbal communication.": Please indicate that this is a self-reported or self-perceived understanding of verbal communication. I think it is important to distinguish the results from objective speech recognition testing (acknowledging that self-reported benefit is very important).
- "...greatest amount of benefit...": Please indicate that it is a self-reported or self-perceived benefit.

Introduction

Page 2: Authors indicate that auditory and vibrotactile information can be unconsciously and naturally integrated in the brain. It would be helpful if the authors could give some description/details of how the integration is hypothesized to occur—how long it takes and what neural/cognitive mechanisms might support it. Even if this is just a hypothesis, it would provide helpful context.

Page 3: "...can help individuals with high frequency hearing loss better understand speech communication..." Please indicate that it can help them self-report or self-perceive better speech communication.

Page 3: The last sentence is too strong. A device can help improve self-reported speech communication without translating to the types of benefits the authors describe in these various situations/environments. It could say something like "the evidence demonstrate the promise of this technology, which if further developed and refined holds promise for..."—something like this. I think it is OK to indicate that these kinds of benefits are possible in the future, but they are not directly supported by the results of this small study. Lots more work is needed.

Methods

Not much detail about the machine learning algorithm is provided. More detail about how it filters background noise (BN) and identifies phonemes would be helpful. How was the algorithm trained? Assuming it was trained on speech, what regional accents were used?

Related to the above, it is not clear how the sham algorithm was used in developing the algorithm. Additional detail/description would be helpful.

Page 4: The authors mention that the algorithm performed poorly for some consonants that people with hearing loss have trouble hearing. It is not clear what level of performance constitutes poor performance and what level constitutes good performance for the phonemes that were selected for the algorithm. More context here would help the reader to understand the results. Understanding the algorithm's accuracy is important for contextualizing the users' results. It would be reasonable to suspect that the users' results should be closely linked to the algorithm's accuracy.

Tasks

I am a bit confused as to why objective speech recognition testing was not completed. The self-reported benefit is absolutely important, but based on the Introduction, the reader is interested in knowing how objective speech recognition improved with the wristband for the selected consonants. If these data are available, it would be helpful to add them. If not, it would be helpful if the authors could explain—somewhere in the manuscript—why this testing was not completed/reported.

Final questionnaire: It does not seem like the results of the final questionnaire are reported in this manuscript. Given that the Abbreviated Profile of Hearing Aid Benefit (APHAB) is the only reported outcome measure, it would be helpful to add these results as well, as they represent something more holistic than the weekly APHAB results.

Paradigm

Does the wristband provide any data logging to indicate how many hours per day the devices were worn? If not, this is not a major flaw but should be mentioned as a limitation because it seems like wear time could directly affect benefit.

APHAB

It might be helpful to the reader to clarify that higher raw APHAB scores indicate worse performance and lower scores indicate better performance but higher benefit scores represent more benefit or better outcomes.

Participants

I would suggest adding the number who did and did not use hearing aids in this section. Any additional information regarding participants—gender, education, etc—would be helpful if it is available to report. Otherwise, I suggest adding that a limitation of the paper is the limited demographic information of the participants (combined with a small n), which makes it hard to determine if any participant-level characteristics might influence the benefit of the wristband.

The authors mention that if a clinical audiogram was unavailable, participants completed an audiogram via a mobile app. Then the authors provide an example, Mimi. Did everyone who used a mobile app use the Mimi app, or did some use other apps?

Relatedly, it would be helpful to report how many participants had clinical audiograms and how many used an app to provide context for the audiometric results.

Results

One critique is that I did not feel like I got a very good handle on the descriptive statistics before the authors started showing group means (with SE of the mean) and comparisons (both over time and between subgroups of participants). I felt that the results emphasized group means (with SE of the mean), but I did not get a good sense of the range and dispersion of the data. In the Discussion, the authors start discussing the numbers of participants who started or ended at a specific APHAB overall score range, but I did not feel like I had the information in the paper to help me contextualize that discussion (because the results, as presented, do not give a very clear view of how individual participants may have performed).

To address the above point, I would strongly suggest adding a descriptive results table first that gives the means, maximums and minimums, and SDs for the overall APHAB scores and, possibly, APHAB benefit scores. It would be nice to see these values for the full participant group, as well as for the subgroups of participants with and without hearing aids (including the n in each group). It would be helpful to see the same data for the subscale scores (ease of communication [EOC], BN, reverberation) if it fits in the table, but I think the overall APHAB scores would be sufficient if space is an issue. Another consideration is that with only 16 participants, you could show the individual-level data for each participant, who would each be a row, and then give the group data in a different row. I defer to the authors on their preferences but would simply suggest that some revisions be made to give the reader a better grasp of the descriptive results.

In the section where subscale analyses are given, the write-up describes comparisons of subscale benefit scores between the different subgroups (with and without hearing aids) as well as comparisons, within a subgroup, of benefit scores to the baseline score. Throughout this section, it is hard to track which *P* values go with which comparisons. It is hard to read and interpret. Additionally, the information is presented slightly differently for each subscale, which makes it even harder to follow. Clearer written descriptions of each comparison being tested—then followed by the statistical numbers—would be beneficial for the reader. Additionally, using a parallel results presentation for each subscale would be helpful.

Discussion

"...individuals with high frequency hearing loss are able to improve their understanding of speech communication...": I would like to see it be specified that this is an improvement in self-reported or self-perceived understanding of speech communication. Previous hearing aid research shows there can

be a placebo effect associated with the perception that one is wearing advanced technology [2].

"...participants were able to improve their ability to understand conversations during daily interactions.": Same comment as above. Please indicate this is a self-reported or self-perceived ability to understand conversations.

"We further found that participants who started the study with a higher APHAB score experienced a greater improvement in their ability to understand speech by the end of the six week trial.": As mentioned earlier in this review, this result is hard to interpret without any sense for the individual variability in the data. The results are presented as group means without clear maximums and minimums or SDs. Providing this information in the Results would help give context to this claim in the Discussion.

"Out of 16 participants, 14 ended the study with an APHAB score of 40 or below...": This is again difficult to interpret without any sense for the individual-level data. At timepoint zero, the group mean is right around 40. It is not clear if ending the study at 40 or below indicates benefit or is just a reflection of peoples' starting points. The discussion should be framed in terms of the amount of benefit people reported.

"Five participants started the study with an unaided APHAB of 50 points or higher...": Again, a better sense of the individual-level data and dispersion would help give context for this. The Results are focused on means and then the Discussion brings up individual data, and it is hard to interpret the two together.

A small point but should this be <30 not >30 as written in the text?

"One potential hypothesis...": It seems like this could also be due to having more room to improve their everyday speech understanding. I think it is important to acknowledge possible noncortical factors that could explain this finding (though I think it is fine to also leave the possibility that it reflects cortical characteristics).

"Participants without hearing aids benefitted the most...": I think given the lack of statistical significance in the comparison of the group means, this needs to be toned down a bit. Perhaps something like "Participants without hearing aids demonstrated a trend toward higher self-reported benefit, though this did not reach statistical significance." I know the authors reference the Cox 10-point criterion, but I am not sure that can be accurately applied to these data when the statistical test says the group means themselves are not statistically different (maybe related to the small sample size and variance in the data). Again, I would also like the benefit to be specified as self-reported or self-perceived.

"Given that this group started the study with a higher APHAB score...": I did not find where there is a statistical test to justify this claim. This should be justified with a *t* test. Otherwise, I think it would be OK to specify that the *t* test did not show a statistical difference, but this group is trending toward having a higher baseline APHAB score.

"In this study, we demonstrated the addition of vibrotactile feedback in the presence of background noise enabled individuals who did not wear hearing aids to hear speech communication better...": Again, would like to see it noted that this is a self-reported or self-perceived benefit.

The authors present the final average BN scores (eg, 28.95 and 40.04), but the section above seems to be focused on benefit scores. This reflects my earlier comment about providing more descriptive data upfront. It is hard to track how the authors switch between baseline and benefit scores, and without a descriptive table to refer to, it is difficult to contextualize some of the Discussion.

Related to the above, these scores are presented as being different but are they statistically different?

"...suggesting that those who use hearing aids may benefit from using vibrotactile feedback during conversations in background noise instead of using their hearing aids.": I think this is much too strong of a conclusion for the data, study design, and sample size. This needs to be significantly toned down—as written, I think this is a reckless conclusion based on the limitations of the data. I could be OK with presenting this as an interesting finding worth future research to determine if the above could potentially be true. However, it would need to be framed by saying that this sort of clinical recommendation would require much larger, more rigorous studies with blinding of participants and researchers.

"Similar to our findings in background noise, we also found...": From what I can see in the subscale results discussion, the difference between the group with and without hearing aids did not reach statistical significance. If this is true, it seems to be going too far to say that the wristband helped people without hearing aids the most. Here, I think it is OK to note that the results are trending in this direction as long as it is acknowledged that the results did not reach statistical significance.

"At the end of the trial, the group of participants who did not wear hearing aids showed an average reverberation score that was less than the average for the group who were regular hearing aid users.": Was this tested statistically? From what I can tell, it looks like only the benefit scores are presented in the Results—not the raw scores. If the Discussion brings up the raw score (not the benefit), this should be presented in the Results section. Again, statistical results are needed to draw conclusions regarding the comparison of means.

"It is possible that individuals who use hearing aids may find haptic vibrations to be more helpful in reverberant environments...": Similar to a comment above, I could be OK with presenting this as an area for future research, but I think it needs to be framed by noting the limitations of this study for drawing any clinical recommendations around hearing aids versus haptic vibration.

"Upon completion of the trial, the average EOC score...": Similar to previous comments, it seems that the Results only present benefit scores but now the Discussion mentions raw EOC scores for the group with and without hearing aids. If raw scores are mentioned in the Discussion, they should be presented in the Results.

This section ends by noting equivalent ending EOC scores for the group with and without hearing aids; a statistical result should be presented to make this claim (and should be presented in the Results section).

One additional note: Results from the final questionnaire do not seem to be presented. Is there a reason for this? Given that the APHAB is the only outcome measure, it would be beneficial to see results from the final questionnaire in this paper alongside the APHAB. The final questionnaire also measures something a little different than the APHAB—it is more holistic for the whole field trial experience.

Conclusion

Same comments as before about noting that this study applies to self-perceived or self-reported benefit.

"We found that vibrotactile feedback provides more benefit for those without hearing aids than for those with hearing aids...": From what I see in the Results section, the statistical results do not support this conclusion. The 10-point criterion from Cox cannot be applied if we are not sure the group means themselves are even different (as indicated by the insignificant *P* value). I think it is OK to say the data are trending in this direction and that the small n may render the study underpowered to detect this difference at $P<.05$. Future work is needed to establish whether this claim is true. For now, I would argue it needs to be softened based on the findings and limitations of the study design.

Finally, I suggest adding a limitations section, which could note limitations around:

- Small n
- Reliance on self-report data without objective speech-testing data
- Potential for placebo effect to influence results
- Small n makes it difficult to discern whether/how individual and demographic characteristics could affect ability to integrate the haptic vibrations and benefit from the wristband—some characteristics one might wonder about include baseline cognitive ability, education level, differences in underlying degree/configuration of hearing loss, or duration of hearing loss
- Use of nonclinical audiogram for some participants (a minor limitation but should be noted)
- No information on how many hours per day the wristband was worn. One might hypothesize that outcomes could be related to wear time. Furthermore—beyond raw wear time—we also do not have information about the richness/complexity of auditory information processed through the wristband

Round 2 Review

I appreciate the authors' thorough revision in response to reviewer feedback, and I found this version to be very much improved. It has been a pleasure reviewing this paper and learning more about the authors' interesting work on this novel device, which is now more clearly and thoroughly explained in this newest version of the paper.

I have only a few suggested minor revisions remaining, as follows:

- In the Results section of the Abstract, it says "those without hearing aids showed a 10.78 point greater drop in average APHAB benefit score at 6 weeks." I believe this should read 10.78 higher APHAB benefit score. It would be a drop in score from baseline to the 6-week score if discussing the global APHAB score, but if discussing the benefit score, then the score increased from baseline to 6 weeks.
- In the Results section of the Abstract, most of the results are discussed as the group average, with only one result framed in terms of non–hearing aid users versus hearing aid users. It might be helpful to more clearly specify that when the average results are presented—it is across all participants. I do not have a strong preference on this, just something I noticed.
- In the last paragraph of the Introduction, it says "...can help individuals with high frequency hearing loss to feel more confident in their ability to understand speech communication." Although I understand why the authors are making this inference from the APHAB, it does not feel quite supported enough to jump from the APHAB results to a statement about participants' confidence. I would strongly suggest editing this to be in line with the language used throughout the rest of the paper (eg, increasing subjective assessment of speech ability, increasing self-rated communication ability, or decreasing self-perceived hearing difficulty in daily communications).
- At the end of the APHAB section under Tasks, where it says "Higher benefit scores indicate...," I would also suggest adding the calculation for the benefit score as unaided – aided; then, it could be deleted from the next section.
- In Table 1, I would suggest adding a column to indicate which participants had a professional hearing test and which used the app option.
- For the Table 2 legend, I would suggest specifying how precision and recall are calculated in terms of true positives, false positives, etc. Additionally, it would be helpful to know how the F_1-score is calculated.
- In the Results section comparing non–hearing aid users to hearing aid users, the sentence about the 10.78-point difference could be made clearer if it specified that the non–hearing aid users had a 10.78-point higher benefit score than the hearing aid users (rather than just saying there is a difference).
- In the same section of the Results, it says "...average APHAB benefit over baseline..."—since the benefit score reflects a reduction in the APHAB score, I would suggest framing benefit not as being "over baseline" but rather "from baseline."
- In the Discussion section, where it says "Out of 16 participants, 14 ended the study with an APHAB score of 40 or below...." I think this would be more helpful if it said how many of them started the study with a score of 40 or below. I do not have a strong preference, however. Now that individual data are presented, it is much easier to contextualize the results.

- In the Discussion section, it says "It is also possible that participants who started the study with a lower APHAB score had more room for improvement." I think this should say a higher APHAB score, as higher scores mean more perceived difficulty.
- In the Conclusion, it mentions that the study was underpowered to detect the difference between hearing aid users and non–hearing aid users at $P<.05$. This is presented for the first time in the Conclusion, which seems out of place. I would suggest first mentioning this in the Limitations section above. It could also be mentioned in the Conclusion, though, because it's an important point—but reading new information in the conclusion was a bit jarring.

Very Minor Comments

- First paragraph under APHAB under Tasks, suggest revising "they are asking" to "they ask"
- In the same section, suggest revising the two instances of "was referring" to "referred"
- For the Results section that discusses the BN score, it should read "16.99 points higher than those with hearing aids" ("with" is missing)

Conflicts of Interest

None declared.

References

1. Kohler I, Perrotta MV, Ferreira T, Eagleman DM. Cross-modal sensory boosting to improve high-frequency hearing loss: device development and validation. JMIRx Med 2024;5:e49969. [doi: 10.2196/49969]
2. Bentler RA, Niebuhr DP, Johnson TA, Flamme GA. Impact of digital labeling on outcome measures. Ear Hear 2003 Jun;24(3):215-224. [doi: 10.1097/01.AUD.0000069228.46916.92] [Medline: 12799543]

Abbreviations

APHAB: Abbreviated Profile of Hearing Aid Benefit
BN: background noise
EOC: ease of communication

Submitted 21.12.23; this is a non–peer-reviewed article;accepted 21.12.23; published 09.02.24.

Please cite as:
Anonymous
Peer Review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"
JMIRx Med 2024;5:e55728
URL: https://xmed.jmir.org/2024/1/e55728
doi:10.2196/55728

Peer Review of “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45973

Companion article: https://www.medrxiv.org/content/10.1101/2023.01.21.23284795v1

Companion article: https://med.jmirx.org/2024/1/e60384

Companion article: https://med.jmirx.org/2024/1/e45973

(JMIRx Med 2024;5:e60428) doi:10.2196/60428

KEYWORDS

cardiac surgery; artificial intelligence; risk prediction; machine learning; operative mortality; data set drift; performance drift; national data set; adult; data; cardiac; surgery; cardiology; heart; risk; prediction; United Kingdom; mortality; performance; model

This is the peer-review report for “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis.”

Round 1 Review

General Comments

Overall, I think this is a really interesting paper [1]. It is a concept I had never heard of, and I can see very clearly how this is an important consideration. I also think the authors have done excellently to consider a host of different aspects, including feature importance change, beyond the most obvious measurements.

Specific Comments

Abstract

1. “It has been suggested that using Machine Learning (ML) techniques, a branch of Artificial intelligence (AI), may improve the accuracy of risk prediction.” Improve them over what? Specify what the status quo is with regard to first principles and data-driven modeling. This statement is also repeated in the first line of the introduction—what is “conventional” about these models?

2. “five ML mortality prediction models”—it should be highlighted that these are novel models that you have developed for this paper.

3. “geometric average results of all metrics”—it is not all metrics, just the 5 that you have calculated. It is better to just say here “a novel metric called the CEM” or something.

Introduction

Why is data set drift a problem? I think you could do more here to highlight how important this is to an audience who might not be dealing with the data themselves and, thus, might not naturally think of examples: for example, changes in treatment guidelines, demographics, new risk factors emerging, or changes in coding practices. You could mention “new” comorbidities such as long COVID.

Methods

1. Could the same individuals be in both the training and validation set and holdout set, if they had multiple surgeries? If so, this may have introduced some bias into the performance estimates. I do not think you need to redo the analyses, but if you can highlight the degree of overlap, then that would be good. Otherwise, say it was not possible and list it as a limitation.

2. “As a sensitivity analysis, we excluded the True Negative Rate from the performance evaluation, by calculating the F1 score.” This sentence does not quite make sense to me. The F_1-score is based on the sensitivity (true negative rate) and the precision (positive predictive value), right? It does not exclude the true negative rate per se; it just does not use it.

Conflicts of Interest

None declared.

Reference

1. Dong T, Sinha S, Zhai B, et al. Performance drift in machine learning models for cardiac surgery risk prediction: retrospective analysis. JMIRx Med 2024;5:e45973. [doi: 10.2196/45973]

Edited by E Meinert; submitted 10.05.24; this is a non–peer-reviewed article;accepted 10.05.24; published 12.06.24.

Please cite as:
Anonymous
Peer Review of "Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis"
JMIRx Med 2024;5:e60428
URL: https://xmed.jmir.org/2024/1/e60428
doi:10.2196/60428

Peer Review of “Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/54611

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.13.23298496v1

Companion article: https://med.jmirx.org/2024/1/e60194

Companion article: https://med.jmirx.org/2024/1/e54611

(JMIRx Med 2024;5:e60394) doi:10.2196/60394

KEYWORDS

human brucellosis; livestock; clustering; spatial; temporal; Iraq

This is the peer-review report for “Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018.”

Round 1 Review

General Comments

This paper [1] talks about human brucellosis in Iraq and brings an interesting spatiotemporal analysis of the human cases in the country. The paper will contribute to the understanding of human brucellosis in Iraq and can be one more example of the use of spatiotemporal analysis for the control of the disease. However, some changes need to be made to clarify some information in the paper.

Specific Comments

Major Comments

Introduction

1. References are missing in the second phrase of the second paragraph.

2. In the third phrase, “(dogs)” is not necessary.

3. The breeding season of sheep and goats is not in spring. So, the phrase “However, it usually coincides with the livestock breeding season, spring” should be changed, as well as “Human exposure to livestock or their contaminated products will occur during spring.” The second part of this phrase is true but not absolute, since animal products can travel to other places; so for clarity, I think this phrase should be reformulated.

4. Please, develop this paragraph further: “A study from northern Iraq showed that the prevalence of brucellosis in livestock varied from 1% to 70%, depending on the species and diagnostic methods. 4 Veterinary vaccination program started in 2007. However, its implementation was negatively affected by insecurities in the region.4.” It would be very interesting to know more about the testing and the insecurities of the population regarding vaccination.

5. Please connect more the idea of the first phrase of the fourth paragraph with the following ideas.

6. Please define “MOH’s” in the last paragraph.

Methods

7. Please make it clear throughout that the data are about humans and not animals.

8. In the data description, why does the data come from different sources and with different types of organization and grouping? Can you make it clearer?

9. What is HB? Please define it before using the abbreviation.

10. What is period one?

11. What does “the low or high attribute zone” mean?

12. Please explain the *P* values assumed for the “Getis-Ord Gi* statistic” and what exactly this statistic is identifying.

13. In the *Abstract*, the statistical analysis is explained differently from the *Methods*. Please make them similar.

Discussion

14. Have you tested the trend? If so, please clarify the methods and results, but if not, please reformulate the phrase where the word is being used to describe the occurrence of your data.

15. Can you define what were the “ISIL events”?

16. “The number of females has been constantly higher than males.” You are talking about humans, please used woman and man and clarify, in the *Methods* section, whether this classification was self-made or not.

17. “housekeeping and farming activities. 5 (11, 12,13).” I did not understand the different configuration of the references in here.

18. “However, this age category was very broad and could have been classified into two to three age groups to detect the most commonly affected age group. (14-17).” Why was this change not made? Again, the references are in a strange configuration.

19. “The infected animals must be eradicated by slaughtering and burning because there is no curable medical therapy for animal brucellosis.” I did not understand this phrase. Please remove the word eradication from here and explain better why you are saying that the animals should be burned. I do understand that positive animals should be slaughter and their carcasses should be disposed of in the right way, but I have not heard of burning before.

20. “On the other hand, humans may consume this infected milk unpasteurized, resulting in infection and areas endemic with brucellosis to animals and humans.” Change “to” for “in.”

21. There is an “18” in paragraph 5 that could be a reference. I did not understand the phrases that followed the 18.

22. The paragraph “Transboundary transfer of animal brucellosis in the region from the neighboring countries such as Iran, Turkey, and Syria were provoked by war and political instability, lack of immunization and animal quarantine, frequent trading, low awareness and poor knowledge of HB prevention and control, residents with poor sanitary conditions easily exposed to Brucella contaminated food and water sources.” is disconnected from the rest of the text; I did not understand what exactly this is about.

23. The first and second periods of the study are not clear for me.

Conclusion

24. The last part of the conclusion would be better in the *Discussion* section, such as “Preventive measures such as health education activities should be performed in high-risk areas. Adopting the Quarantine-Slaughter-Immunization strategy and One Health Approach is crucial in controlling the disease. This can be achieved through multisectoral coordination and coordination with neighboring countries in the control programs.”

Round 2 Review

General Comments

This paper brings important information and analysis on human brucellosis in Iraq. To improve the paper’s understanding, I suggest an English review of the paper to improve the writing of the paper.

Specific Comments

Major Comments

1. *Abstract*, section *Methods*: “The trend of cases by sex and age group were displayed from 2007 - 2018 were displayed.” Please delete the last “were displayed.”

2. “The seasonal distribution of the cases from 2007 to 2012 was graphed.” Substitute “was” for “were.”

3. *Introduction*: The paragraph on the percentages of occurrence of brucellosis only present the values but does not make a value judgment or interpret what these values mean, why are they important, and so on. Please, reformulate again.

4. *Discussion*: Second paragraph: Make it clear that the number of woman you are talking about is the number of woman positive for brucellosis among the analyzed years.

5. *Conclusion*: Substitute HB for human brucellosis.

Conflicts of Interest

None declared.

Reference

1. Mustafa AH, Khaleel HA, Lami F. Human brucellosis in Iraq: spatiotemporal data analysis from 2007-2018. JMIRx Med 2024;5:e54611. [doi: 10.2196/54611]

Edited by E Meinert; submitted 09.05.24; this is a non–peer-reviewed article;accepted 09.05.24; published 03.07.24.

Please cite as:
Anonymous
Peer Review of “Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018”
JMIRx Med 2024;5:e60394
URL: https://xmed.jmir.org/2024/1/e60394
doi:10.2196/60394

Peer Review of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial"

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.14.23292603v1

Companion article: https://med.jmirx.org/2024/1/e56442

Companion article: https://med.jmirx.org/2024/1/e50970

(JMIRx Med 2024;5:e57310) doi:10.2196/57310

KEYWORDS

leprosy; ulcers; wounds; honey; neuropathy; nerves; Africa; randomized controlled trial; RCT

This is the peer-review report for "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial."

Round 1 Review

General Comments

This paper is a protocol description of an important study [1], especially for contexts in which advanced wound care products are often not available. It is a well-written protocol with clear steps to take. Below are some of my feedback; I also included some small textual feedback points in the text. You may not be able to address all the points I raised, as it seems that the trial has already started, but in that case, it would be interesting to describe why or why not in the manuscript's text.

Specific Comments

Major Comments

1. Please describe why the 84-day cutoff period was chosen.

2. The flowchart is a bit small and thus hard to read.

3. Usually, overlapping inclusion and exclusion criteria are not mentioned.

4. It is not clear why hepatitis B or C were added in the exclusion criteria list.

5. It is not clear for me if patients are clinically admitted or not, and if so, why? For how long? Is this routine care? And what are the discharge criteria?

6. If diabetes is excluded, it may be good to also exclude other known reasons for peripheral neuropathy (eg, vitamin B deficiencies)

7. Are signs of infection also monitored, assessed, or outcome measures?

8. Please explain more on the swabs: what kind of swab is it and what is tested, if this is not part of the research project? In general, it is better to take a routine swab to test for infection (bacterial growth) prior to inclusion instead of prior to randomization, as infection is an exclusion criteria. Also, address this under the heading about "biological specimens."

9. Explain why the video recording is taking place. It may be interesting to also do it with the last 5 patients if it is performed for monitoring reasons.

10. Why are assessors from Nepal used and not contextual assessors from Nigeria itself (also, is it because of skin color differences of participants in both countries)?

11. Please explain why early analysis is taking place after the inclusion of the first three-eighths of participants.

12. I missed the argument in the discussion that mentioned that honey is often relatively cheap and better available than many advanced wound care products.

13. Include some information about how long data will be stored (number of years), where it will be stored in a secure way, and if it will be shared (pseudonymized) if requested (eg, for reproducibility).

Minor Comments

14. Write numbers up to 9 in text.

15. Check abbreviations.

16. Update reference list, include authors, URLs and "assessed on [date]" in references to websites and online documents.

17. Explain the camera used for photography.

18. Please add 2 more references in the discussion.

19. Explain more about the pedometer usage.

Conflicts of Interest

None declared.

Reference

1. Udo S, Ogbu Sunday P, Tsaku PA, et al. Raw, unadulterated African honey for ulcer healing in leprosy: protocol for the Honey Experiment on Leprosy Ulcer (HELP) randomized controlled trial. JMIRx Med 2024;5:e50970. [doi: 10.2196/50970]

Edited by E Meinert; submitted 12.02.24; this is a non–peer-reviewed article;accepted 12.02.24; published 01.03.24.

Please cite as:
Anonymous
Peer Review of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial"
JMIRx Med 2024;5:e57310
URL: https://xmed.jmir.org/2024/1/e57310
doi:10.2196/57310

Peer Review of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence"

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/51787

Companion article: https://med.jmirx.org/2024/1/e55899

Companion article: https://med.jmirx.org/2024/1/e51787

(JMIRx Med 2024;5:e56047) doi:10.2196/56047

KEYWORDS

animal-assisted therapy; pet therapy; outcome assessment; policies; systematic study

This is a peer-review report submitted for the paper "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence."

Round 1 Review

General Comments

I enjoyed reading this paper [1]. In general, this is a well-written paper. There are some areas that could be clarified or expanded to improve the strength of the article. Due to the justification, the spacing of punctuation marks appears incorrect, and there are rare instances of double punctuation (periods). The use of American Psychological Association abbreviations at first use was not followed. At times, after providing an abbreviation, the full name is spelled out (eg, "AAT," "PTSD").

Specific Comments

Major Comments

1. There does not appear to be a Table 2, but it is referenced in the text (page 7).

2. In the Discussion section on page 15, a reference is made to effect sizes in four outcome areas, yet no effect sizes were reviewed in the article.

3. Also in the Discussion on page 15, the word "power" is used: "The increased number of studies provided greater power in assessing variance heterogeneity and potential group differences." Unless a specific power analysis was performed (if so, it should be discussed), the word "power" could be changed to "support" to reflect a review rather than an analysis. Similarly, on page 16, the term "meta-analysis" is used. Unless a secondary analysis of pooled data was performed, the term "meta-analysis" should be changed to "analysis." If a secondary pooled analysis was performed, that should be defined and described in the body of the paper.

4. Page 16, Society for Healthcare Epidemiology of America is noted as an organization of interest for animal-assisted therapy, as is Pet Partners. The authors may want to consider including the global organization called International Association of Human-Animal Interaction.

5. On page 17, the authors cite lack of blinding as a limitation and introduction of bias. I would be curious to know how the authors propose blinding in studies that involve interactions with animals. I strongly suggest this sentence be removed.

6. The work done by Hinic and others [2] was not a randomized controlled study as noted in Table 1. Please double-check that all studies listed are correctly labeled as randomized studies.

Minor Comments

7. Check punctuation for spacing.

8. Check all abbreviations and use abbreviations after first use is defined.

9. Check capitalizations in midsentence (page 4: "Dog"; page 7: "Unrepresentative").

10. Page 6: "The articles should to be published in English." Wrong tense—change to "were."

11. Page 1: Three categories of interventions were provided in section 2.4. It would strengthen the paper to include definitions of these categories for the reader.

Conflicts of Interest

None declared.

References

1. Pandey RP, Himanshu, Gunjan, Mukherjee R, Chang C. The role of animal-assisted therapy in enhancing patients' well-being: systematic study of the qualitative and quantitative evidence. JMIRx Med 2024;5:e51787. [doi: 10.2196/51787]
2. Hinic K, Kowalski MO, Holtzman K, Mobus K. The effect of a pet therapy and comparison intervention on anxiety in hospitalized children. J Pediatr Nurs May-Jun 2019;46:55-61. [doi: 10.1016/j.pedn.2019.03.003] [Medline: 30852256]

Edited by E Meinert; submitted 03.01.24; this is a non–peer-reviewed article;accepted 03.01.24; published 18.03.24.

Please cite as:
Anonymous
Peer Review of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence"
JMIRx Med 2024;5:e56047
URL: https://xmed.jmir.org/2024/1/e56047
doi:10.2196/56047

Peer Review of “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis”

Juntong Zeng, MD, PhD

Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45973

Companion article: https://www.medrxiv.org/content/10.1101/2023.01.21.23284795v1

Companion article: https://med.jmirx.org/2024/1/e60384

Companion article: https://med.jmirx.org/2024/1/e45973

(JMIRx Med 2024;5:e60280) doi:10.2196/60280

KEYWORDS

cardiac surgery; artificial intelligence; risk prediction; machine learning; operative mortality; data set drift; performance drift; national data set; adult; data; cardiac; surgery; cardiology; heart; risk; prediction; United Kingdom; mortality; performance; model

This is the peer-review report for “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis.”

Round 1 Review

General Comments

This manuscript [1] presents an interesting study that explores temporal trends in various performance metrics for different types of prediction models used in the prediction of in-hospital mortality after cardiac surgery in the United Kingdom from 2012 to 2019. The data set was divided into 2 periods: from 2012 to 2016 for model training and internal validation and from 2017 to 2019 for external validation. The study evaluated 5 prediction models: logistic regression, support vector machine (SVM), random forest, extreme gradient boosting (XGBoost), neural network, and European System for Cardiac Operative Risk Evaluation (EuroSCORE) II. The authors aimed to assess the model performance on 5 metrics (1 – expected calibration error [ECE], area under the curve [AUC], 1 – Brier score, F_1-score, and net benefit) and proposed a composite metric, the clinical effectiveness metric (CEM), calculated as the geometric mean of the 5 mentioned metrics, as the primary metric.

The study began with a nontemporal baseline evaluation of different models in the 2017 - 2019 temporal validation and then conducted a series of drift analyses, including an examination of overall trends from 2012 to 2019, within-period trends in the first 3 months of 2017 and 2019, and between-period trends between the first 3 months of 2017 and 2019. The authors also analyzed drift in variable importance and variable distribution, defined by the temporal change in the ratio of several top-importance features within the data set, to profile data set drift.

The authors demonstrated that XGBoost and random forest were the best-performing models, both in nontemporal and temporal evaluations, whereas the EuroSCORE II model exhibited a significant drop in performance. Temporal declines in model performance were observed across all models and were consistent with data set drift.

Overall, the question of the generalizability of prediction models, whether temporal or spatial, has long been a topic of discussion in clinical research. This study takes a commendable approach to addressing this question. However, there are some issues that require clarification and revision, including (1) methodological concerns related to the justification of the main metric (CEM) using averaging, and the appropriateness of some statistical tests; (2) the clinical significance of the identified performance drift; and (3) the overall clarity of the study’s design and presentation.

Specific Comments

Major Comments

1. The statement of the study’s objectives should be improved for more clarity, particularly regarding the phrase “verify suspected dataset drift by assessing the relationship between and within performance drift, variable importance drift, and dataset drift across ML and ES II approaches.” It is unclear what is meant by the “relationship between and within.” Does this refer to the analysis of performance drift within and between different periods? The overall study design is quite challenging to grasp initially, even with the graphical overview provided in Figure 1. To enhance clarity, additional details and explanations

should be added to the aims, overall design, graphical overview, and text the *Methods* and *Results* sections.

2. The rationale for introducing CEM as the primary performance metric, calculated as the geometric mean of 5 distinct individual metrics, is debatable and lacks strong justification. Although the geometric mean is less sensitive to outliers compared to the arithmetic mean, it raises the fundamental question of why these metrics need to be summarized. Is it merely to obtain a single quantitative measure for analysis, or does it aim to provide a more comprehensive understanding of overall model performance? It appears to serve primarily the former purpose, which may not be an appropriate practice given that the 5 metrics assess entirely different aspects of model performance: 1 – ECE for calibration, AUC for discrimination, 1 – Brier score (which already encompasses calibration and discrimination components), F_1-score for threshold-specific discrimination, and net benefit index for cost-effectiveness. Consequently, interpreting the exact meaning of CEM becomes challenging, as it reduces these diverse aspects to a single numerical value. Therefore, I suggest just reporting and examining all 5 metrics individually, with or without highlighting certain ones as primary areas of interest.

3. The manuscript used several statistical tests, and some of them are relatively less commonly used. Please provide a more detailed description of the objectives and specific statistical situations for each test used. Additionally, for the baseline nontemporal performance comparison, a more conventional approach for comparing AUC would be the use of the DeLong method (you could choose the best model as the reference), and bootstrapping can be used to assess the statistical significance when comparing other metrics.

4. During the training and internal validation phase with 5-fold cross-validation, additional details are needed to understand how the final model for each model type was selected for subsequent temporal validation, including whether hyperparameter tuning was carried out and whether there was a final refitting process on the entire training data set following the cross-validation, etc.

5. The *Introduction* section should incorporate more background information on previous studies reporting or relating to performance variation in prediction models for cardiac surgery outcomes. In the *Discussion* section, it is also important to discuss how this work contributes to existing evidence in the context of these previous studies. Some relevant studies, based on my preliminary search, include Benedetto et al [2], Zeng et al [3], Mori et al [4], and potentially more.

6. Although the authors observed numerical declines in CEM and other metrics, the magnitude of these declines appears to be relatively small, particularly when considering metrics such as AUC. As a result, it is essential to discuss how to interpret this magnitude of drift in the context of clinical practice. In other words, what is the clinical significance of this variation in performance, and how does it justify the necessity of actively monitoring model drift in terms of cost-effectiveness? Please discuss.

7. The conclusion should only focus on the primary findings outlined in the aims of the *Introduction* section. Avoid incorporating less central findings and speculative elements. Additionally, it may not be fair to suggest replacing the EuroSCORE II model simply based on the inferior performance in this study, since it was already established and this study essentially conducted an external validation for it, whereas the other machine learning models were developed using these data sets.

Minor Comments

1. More detailed definitions and explanations should be provided for each performance metric.

2. In the *Methods* section, please provide a clear outline of the inclusion and exclusion criteria. Additionally, consider including a flowchart that illustrates the data set development process, outlining how these criteria were applied.

3. I had difficulty understanding what “outliers” and “distribution” meant in the *Results* section for the baseline nontemporal performance of each model. I thought that each metric of each model should be just a numerical value and a 95% CI from bootstrapping.

4. The title of the manuscript should be an objective reflection of the overall study design and aim, rather than drawing conclusions from the findings.

5. I did not find the supplementary materials in the review system. I am not sure whether this issue is on my end or not.

Round 2 Review

General Comments

I appreciate the opportunity to rereview this manuscript. The authors’ efforts in revising their manuscript in response to previous concerns are commendable. This manuscript has been improved and is now in principle publishable. It could potentially be accepted upon reasonable response to a few follow-up minor comments, outlined below.

Specific Comments

1. About my previous major comment 1, the authors meticulously elaborated on (1) the reasons for performance drift and (2) its importance, which are both valid points. However, the current *Introduction* (lines 121 - 179) is quite lengthy. I recommend consolidating these 2 parts into a single paragraph, listing each point without the need for detailed individual explanations. Additionally, my query about the exact meaning of “the relationship between and within variable importance drift, performance drift, and actual dataset drift” remains unaddressed. Even though it was removed from the *Introduction*, it still appears in the abstract. I suggest the authors explicitly explain it to readers and incorporate it into the manuscript when first mentioned.

2. Regarding the justification for the CEM, the authors have added more explanation and supporting literature for its use. However, it would strengthen their case if they could provide examples from external studies or use cases where a similar

practice (averaging different aspects of metrics for model performance evaluation) was used, beyond their own studies.

3. About the statistical tests for comparing AUC with the DeLong method, I believe that performing the DeLong test for AUC comparison is not overly computationally demanding, even on a relatively large data set. I recommend the authors explore commonly used R packages (eg, “pROC”) that facilitate AUC calculation and comparison with the DeLong method. The DeLong comparison typically requires paired variables of the label and 2 models’ predicted probabilities, and the 95% CI and *P* value are automatically calculated by bootstrapping these paired samples, which is relatively efficient.

4. Regarding model tuning and specification of the best models (PS: I still cannot find the supplements, only a revised clean manuscript; I am not sure if this was due to issues from my end), I am curious why different tuning practices were used for different models, especially grid search for XGBoost and SVM but manual tuning for random forest.

5. In response to the query about the clinical significance of the relatively small scale of performance drift, the authors referred to one of their previous studies briefly discussing this matter. However, it would be much clearer if the authors could more explicitly elaborate in this study and, if possible, provide additional analysis to support this argument.

Conflicts of Interest

None declared.

References

1. Dong T, Sinha S, Zhai B, et al. Performance drift in machine learning models for cardiac surgery risk prediction: retrospective analysis. JMIRx Med 2024;5:e45973. [doi: 10.2196/45973]
2. Benedetto U, Sinha S, Lyon M, et al. Can machine learning improve mortality prediction following cardiac surgery? Eur J Cardiothorac Surg 2020 Dec 1;58(6):1130-1136. [doi: 10.1093/ejcts/ezaa229] [Medline: 32810233]
3. Zeng J, Zhang D, Lin S, et al. Comparative analysis of machine learning vs. traditional modeling approaches for predicting in-hospital mortality after cardiac surgery: temporal and spatial external validation based on a nationwide cardiac surgery registry. Eur Heart J Qual Care Clin Outcomes 2024 Mar 1;10(2):121-131. [doi: 10.1093/ehjqcco/qcad028] [Medline: 37218710]
4. Mori M, Durant TJS, Huang C, et al. Toward dynamic risk prediction of outcomes after coronary artery bypass graft: improving risk prediction with intraoperative events using gradient boosting. Circ Cardiovasc Qual Outcomes 2021 Jun;14(6):e007363. [doi: 10.1161/CIRCOUTCOMES.120.007363] [Medline: 34078100]

Abbreviations

AUC: area under the curve
CEM: clinical effectiveness metric
ECE: expected calibration error
EuroSCORE: European System for Cardiac Operative Risk Evaluation
SVM: support vector machine
XGBoost: extreme gradient boosting

Edited by E Meinert; submitted 06.05.24; this is a non–peer-reviewed article;accepted 06.05.24; published 12.06.24.

Please cite as:
Zeng J
Peer Review of “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis”
JMIRx Med 2024;5:e60280
URL: https://xmed.jmir.org/2024/1/e60280
doi:10.2196/60280

Peer Review of “Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/54611

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.13.23298496v1

Companion article: https://med.jmirx.org/2024/1/e60194

Companion article: https://med.jmirx.org/2024/1/e54611

(JMIRx Med 2024;5:e60433) doi:10.2196/60433

KEYWORDS

human brucellosis; livestock; clustering; spatial; temporal; Iraq

This is the peer-review report for “Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018.”

Round 1 Review

General Comments

This paper [1] presents a spatiotemporal distribution analysis of the outbreak of the brucellosis in Iraq from 2007 to 2018, providing explanations for potential underlying causes. The methods employed include descriptive analysis and Getis-Ord G_i*. The paper exhibits a well-structured format, clear language, rich content, and appropriate methodology.

Specific Comments

Major Comments

1. The *Abstract* and the main text exhibit inconsistency in describing the methods employed. The *Results* section of the main text only includes the results of the descriptive analysis and Getis-Ord G_i*, with no mention of the Moran *I* method as indicated in the *Abstract*.

2. The methods used in the paper should be briefly explained in the *Method*s section to clarify their principles.

3. In the *Results* section, the authors state that there is an increasing trend in female cases from 2016 onward. This conclusion cannot be drawn; female cases increased from 2016 to 2017 and then decreased by 2018, falling below the 2016 quantity.

4. Include spatial distribution maps of the incidence rates for 1-2 years during the study period.

Minor Comments

5. Add numerical labels to the bars in Figure 1 for a more intuitive understanding.

6. Figure 4 lacks coordinate axes, and there is an incomplete gray box on the horizontal axis, affecting aesthetics.

7. Please provide the formula for calculating the case frequency.

Round 2 Review

Specific Comments

Major Comments

1. Maybe I did not express it clearly, but for the local Getis-Ord G_i* method, which is one of the main methods applied in this paper, the authors should give the formula for its calculation and add the source.

2. This is not a comment that has to be revised. Generally, the significance and spatial location of clusters in the local Getis-Ord G_i* results are shown on the same map; for example, hot spots with different levels of significance are represented by 3 progressively deeper red colors, and cold spots with different levels of significance are represented by 3 progressively deeper blue colors. Also, Figure 5 contains too many maps, and it is more concise to show the results for 1 year in 1 map.

3. The elements that are really necessary inside a map, including but not limited to a scale, a compass, and preferably the addition of national boundaries, are missing.

Other Comments

The authors have finished revising, and I do not have any questions.

Conflicts of Interest

None declared.

Reference

1. Mustafa AH, Khaleel HA, Lami F. Human brucellosis in Iraq: spatiotemporal data analysis from 2007-2018. JMIRx Med 2024;5:e54611. [doi: 10.2196/54611]

Edited by E Meinert; submitted 10.05.24; this is a non–peer-reviewed article;accepted 10.05.24; published 03.07.24.

Please cite as:
Anonymous
Peer Review of "Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018"
JMIRx Med 2024;5:e60433
URL: https://xmed.jmir.org/2024/1/e60433
doi:10.2196/60433

Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"

Felicianus Pereira, DPT

Dow University of Health Sciences, Karachi, Pakistan

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45688

Companion article: https://med.jmirx.org/2024/1/e55863

Companion article: https://med.jmirx.org/2024/1/e45688

(JMIRx Med 2024;5:e56169) doi:10.2196/56169

KEYWORDS

COVID-19; injury; prevalence; adaptation; sports medicine; follow-up; training; football; epidemiology; sport; athlete; athletic; injuries

This is the peer-review report for "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study."

Round 1 Review

General Comments

I would like to commend the authors on performing a follow-up study [1]. This well-executed study provides a good comparison with how COVID-19 disrupted global schedules and the impact it had on the sports sector. The previous version of this study identified how a lack of training greatly increased chances of injuries in National Football League (NFL) athletes. In the current study, the authors have identified how a timely training program can reduce the injury time of athletes. A few points require clarification, such as:

1. What is the starting month of the NFL season? The COVID-19 lockdown was implemented from late March 2020 to late May 2020 (mentioned in the current study). Did this fall at the start, at the middle, or toward the end of the training phase of the athletes?

2. Were athletes provided any equipment at home, or were they recommended any training protocol by their team's coaches (the way some clubs in the English Premier League provided gym equipment to their players at home to maintain fitness or had online practice sessions with their players)?

3. When lockdown restrictions were lifted, how much time were the athletes provided to restore their match fitness?

4. How long does it take for detraining to set in, and how quickly can athletes regain their lost fitness?

Minor Comments

5. Change the tense of some sentences. In the introduction, use "We hypothesized" instead of "we hypothesize" and change the next line to "injury prevalence for the 2021 and 2022 seasons WOULD be lower than the 2020 NFL season..."

Conflicts of Interest

None declared.

Reference

1. Puga TB, Schafer J, Thiel G, et al. COVID-19 National Football League (NFL) injury analysis: follow-up study. JMIRx Med 2024;5:e45688. [doi: 10.2196/45688]

Abbreviations

NFL: National Football League

Edited by E Meinert; submitted 08.01.24; this is a non–peer-reviewed article;accepted 08.01.24; published 13.02.24.

Please cite as:
Pereira F
Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"
JMIRx Med 2024;5:e56169
URL: https://xmed.jmir.org/2024/1/e56169
doi:10.2196/56169

Peer Review of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial"

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.14.23292603v1

Companion article: https://med.jmirx.org/2024/1/e56442

Companion article: https://med.jmirx.org/2024/1/e50970

(JMIRx Med 2024;5:e56498) doi:10.2196/56498

KEYWORDS

leprosy; ulcers; wounds; honey; neuropathy; nerves; Africa; randomized controlled trial; RCT

This is the peer-review report for "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial."

Round 1 Review

General Comments

This is an excellent interventional protocol for a randomized controlled trial assessing honey as a potential ulcer therapeutic [1]. Careful consideration has been made to avoid bias and ensure robust results. I would suggest a few things to consider (below) prior to publishing.

Specific Comments

Major Comments

Background and Rationale

- It is mentioned that 30% to 50% of people infected with leprosy have nerve damage. Be more specific here with "people"—is this a global estimate, American estimate, Nigerian estimate, etc?
- The background may benefit from a more specific discussion of previous literature. If there is a significant systematic review on the topic, a quick summary of relevant findings in the background (or later in the discussion) would help to situate the rationale behind carrying out such a study.

Study Setting

- St. Benedicts Tuberculosis and Rehab Hospital is owned by the Catholic Diocese. Do the authors suspect a potential religious bias in individuals who attend this hospital? Does this affect other social demographics and potentially skew generalizability?
- I am not sure it is necessary to go into this much detail about the staffing compositions and facilities of each site. Consider truncating.

Additional Consent Provisions

- This section mentions that the computer program will range-check information. Please specify which computer program.

Intervention Description

- The honey is being obtained from local bee farmers in North Central Nigeria; however, it is unclear how the honey is being prepared prior to inclusion into the study. I understand it is being checked for botulism (which is great); however—I wonder—is the honey from different farms being mixed together prior to use? Or is it possible that 1 dressing may be from 1 specific farm, etc? If so, is there a potential risk of interventional procurement bias? Meaning the honey from one farm may be better at wound healing then the honey from another farm? Just something to think about…
- Be more clear about the function of the video recording. Will it also be used to test if assessors can distinguish between honey versus control?

Confidentiality

- This section mentions that study forms containing personal identifier information will be kept secured and locked at trial site. Which trial site? Just 1? Or both? Please be more specific here.

Other

- Will you be collecting demographic data such as sex, gender, creed, socioeconomic status, level of schooling, etc? Would you consider stratifying results by any of these parameters?
- Additionally, will you be collecting information on leprosy status, that is, paucibacillary versus multibacillary leprosy or if the patient has progressed into the leprosy reaction stage (type 1 or type 2)? This information may be useful

for downstream analysis and can be stratified for to avoid spectrum bias.

Minor Comments

Background and Rationale

- Edit sentence to read: "…and peripheral nerves, causing neuropathy and severe disability, consequently resulting in social exclusion and stigmatization."
- Edit sentence to read: "…new child cases [3], with a grade 2 disability rate of about 15% for the past…"
- Edit sentence to read: "Thirty to fifty percent of…"
- Edit sentence to read: "Ulcers usually occur in anesthetic feet, and will heal slowly with routine therapy, however have a tendency to recur [6]."
- Edit sentence to read: "…documented report record in the Edwin Smith Papyrus…"
- Edit sentence to read: "…gathered and modified by the honeybee…"
- Edit sentence to read: "…exudates, and possesses antimicrobial…"
- Edit sentence to read: "…treatment of difference kinds of wounds, as researchers continue…"
- Edit sentence to read: "…sizeable number of reports that show mixed levels…"
- Edit sentence to read: "…with only about 5% of patients reporting pain following dressing."
- The rest of the previous sentence "…and undocumented concern of botulism disease due to infection…" is unclear. Do the same 5% of patients also report concerns of botulism? Or is botulism a concern the authors have, and that has not been reported in previous literature? Either way I would make the botulism argument its own separate sentence that is more clear.

Study Setting

- Is it "St. Benedict's TBL" or "St. Benedit's TBL"? Please correct all instances to 1 or the other. The first sentence under study setting uses "St Benedits."
- In "…is a TB and leprosy…" please type out "Tuberculosis" on first use with "(TB)" in quotes as per other abbreviations.

Eligibility Criteria

- Edit sentence to read: "…in the intervention group, all ulcers – not just the one…"
- Edit sentence to read: "Routine swabs will be taken, but the interpretation of…"

Additional Consent Provisions

- Edit sentence to read: "Photograph of the ulcers will be…"

Relevant Concomitant Care

- Edit sentence to read: "…bearing and the level of activity of patients might…"

Recruitment

- Edit sentence to read: "…identified by the on-site clinical…"

Plans for Assessment

- Edit sentence to read: "…database managers at the University of…"

Composition of the Data

- Edit sentence to read: "…Monitoring Committee consists of individuals…"
- Edit sentence to read: "…participant has been followed up for 84 days or discharged, whichever…"

Dissemination Plans

- Low- and middle-income countries needs to be fully written out at first use, then the abbreviation "LMIC" can follow.

Discussion

- Edit sentence to read: "…of its near absence from the global health agenda [25], and as such, very little…"
- Edit sentence to read: "A Cochrane review [31] noted that previously published evidence is limited, due to a high or unclear risk of bias (selection, performance, detection, or attrition) detected, imprecision due to little participants, indirectness due to poor outcome measures, and inapplicable interventions."
- Edit sentence to read: "Although honey has been known for centuries to promoted wound healing, there are only a few controlled clinical trials that assess its efficacy."
- This would be a good place to include a brief discussion of relevant findings with specific outcomes or statistics, as I mentioned in the background section.

Conflicts of Interest

None declared.

Reference

1. Udo S, Ogbu Sunday P, Tsaku PA, et al. Raw, unadulterated African honey for ulcer healing in leprosy: protocol for the Honey Experiment on Leprosy Ulcer (HELP) randomized controlled trial. JMIRx Med 2024;5:e50970. [doi: 10.2196/50970]

Edited by E Meinert; submitted 17.01.24; this is a non–peer-reviewed article;accepted 17.01.24; published 01.03.24.

Please cite as:
Anonymous
Peer Review of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial"
JMIRx Med 2024;5:e56498
URL: https://xmed.jmir.org/2024/1/e56498
doi:10.2196/56498

Peer Review of “Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.08.14.23294073v1

Companion article: https://med.jmirx.org/2024/1/e57116

Companion article: https://med.jmirx.org/2024/1/e52198

(JMIRx Med 2024;5:e57159) doi:10.2196/57159

KEYWORDS

TRICARE; health care fraud; Defense Health Agency; fraud; fraudulent; insurance; coverage; beneficiary; beneficiaries; background check; background checks; demographic; security clearance; FDA; Medicaid; Medicare; provider; provider referral; military; False Claims Act; HIPAA breach; OIG-LEIE; inspector general; misconduct; insider threat; information system; zero trust; data management; Food and Drug Administration; Health Insurance Portability and Accountability Act breach; Office of Inspector General's List of Excluded Individuals and Entities

This is the peer-review report for “Insider Threats to the Military Health System: A Systematic Background Check on TRICARE West Providers.”

Round 1 Review

General Comments

This paper [1] examines the list of TRICARE providers eligible to deliver telehealth whose names appear on one or more federal sanction lists. This work could have a high impact with implications for national security and patient safety. However, it is not well organized and does not seem to adhere to the scientific method.

Specific Comments

Major Comments

1. This is important work, but is it actually science? A team compared two lists. There are no statistics, minimal numbers, and only one hard conclusion (improper monies). Lots of speculation, but no real answers. Because you are calling out the Defense Health Agency (DHA) and the Military Health System (MHS) on inadequate oversight, your conclusions must be driven by airtight methodology and presented in a professional and well-organized manner. Otherwise, you may just submit this work to the DHA’s Office of Inspector General as you’ve already done and call it completed.
2. Assuming you decide to go with the science, this article has important things to say, but it is not yet ready for publication. It is poorly organized, somewhat informal in tone, and comes across as inflammatory in places. An example of poor organization is the focus on cyber threats and potential in the Introduction, improper monies paid to sanctioned providers in the Results, and a distrust of provider data in the Discussion. Patient safety is not discussed until page 17. Recommend mentioning all these issues in the Introduction and then addressing them in the Results/Discussion in a systematic, organized fashion. Also, streamline areas where the same data is repeated multiple times.
3. Similarly, I recommend keeping all the DHA/MHS recommendations together and at the end of the Discussion section, and addressing these in a systematic and organized fashion. “Based on these results, the DHA/MHS/ TRICARE/ whoever should consider the following: (1) Recommendation 1. (2) Recommendation 2...” etc. (Don’t need to take my wording but this is a general idea.)
4. If there are other people who participated in this study, they should be included as authors or in the Acknowledgments. I doubt that one person compared tens of thousands of names solo.

Minor Comments

Everything else is gravy until these major issues are fixed. That said, I sincerely wish you the best of luck as you continue working on this important issue.

Conflicts of Interest

None declared.

Reference

1. Bychkov D. Insider threats to the military health system: a systematic background check of TRICARE West providers. JMIRx Med 2024;5:e52198. [doi: 10.2196/52198]

Abbreviations

DHA: Defense Health Agency
MHS: Military Health System

Edited by E Meinert; submitted 06.02.24; this is a non–peer-reviewed article;accepted 06.02.24; published 09.04.24.

Please cite as:
Anonymous
Peer Review of "Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers"
JMIRx Med 2024;5:e57159
URL: https://xmed.jmir.org/2024/1/e57159
doi:10.2196/57159

Peer Review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"

Vetriselvan Subramaniyan, PhD

Pharmacology Unit, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Sunway City, Malaysia

Related Articles:

Companion article: https://arxiv.org/abs/2304.10512

Companion article: https://med.jmirx.org/2024/1/e57838

Companion article: https://med.jmirx.org/2024/1/e48519

(JMIRx Med 2024;5:e58317) doi:10.2196/58317

KEYWORDS

opioid; substance use; substance use disorder; social media; US; opioid crisis; mental health; substance misuse; crypto; dark web; users; user perception; fentanyl; synthetic opioids; United States

This is the peer-review report for "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study."

Round 1 Review

The paper [1] is titled ""Can We Detect Substance Use Disorder?": Knowledge and Time Aware Classification on Social Media from Darkweb."

General Comments

1. The paper has been written comprehensively.
2. The study aims to analyze substance use posts on social media from the dark web and detect substance use disorder using appropriately developed state-of-the-art deep learning and knowledge-aware Bidirectional Encoder Representations From Transformers–based models.
3. Topic analysis is performed to appropriately identify correlations between different drugs and the topics discussed in social media posts.
4. The most effective model achieves statistically significant performance (macro–F_1-score 82.12, recall 83.58) in accurately identifying substance use disorder.

Minor Comments

1. The study acknowledges the challenges of crawling crypto markets and the restricted crawling process, which limits the data available for analysis.

Need to explain in the manuscript.

2. The study proposes building an opioid drug social media knowledge graph but does not provide details on the potential impact or implications of such a graph.

Need to provide the details in the manuscript.

3. The study explicitly states that it does not make any clinical diagnosis or treatment suggestions, which indicates a gap in translating the research findings into practical applications for addressing substance use disorder.

Need to justify how this study will be helpful for clinical situations.

Report

After incorporating the suggested comments, this paper is suitable for publication in the *Journal of Medical Internet Research*.

Conflicts of Interest

None declared.

Reference

1. Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A. Detecting substance use disorder using social media data and the dark web: time- and knowledge-aware study. JMIRx Med 2024;5:e48519. [doi: 10.2196/48519]

Edited by E Meinert; submitted 12.03.24; this is a non–peer-reviewed article;accepted 12.03.24; published 01.05.24.

Please cite as:
Subramaniyan V
Peer Review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"
JMIRx Med 2024;5:e58317
URL: https://xmed.jmir.org/2024/1/e58317
doi:10.2196/58317

Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"

Ari Samaranayaka, BSc, MPhil, PhD

University of Otago, Dunedin, New Zealand

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45688

Companion article: https://med.jmirx.org/2024/1/e55863

Companion article: https://med.jmirx.org/2024/1/e45688

(JMIRx Med 2024;5:e55997) doi:10.2196/55997

KEYWORDS

COVID-19; injury; prevalence; adaptation; sports medicine; follow-up; training; football; epidemiology; sport; athlete; athletic; injuries

This is the peer-review report for "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study."

Round 1 Review

General Comments

My Review—COVID-19 NFL Injury Prevalence Analysis, A Follow-Up Study

Throughout the manuscript [1], including in the title, the authors say they analyzed the prevalence of injuries. This is incorrect. They have not analyzed injury prevalence; they did not even collect the data required for such an analysis. Instead, they collected injury incidence data and analyzed them.

The primary component of this study is analyzing publicly available data to make a conclusion. I have a major concern regarding the statistical analysis the authors have performed. They have collected injury incidence data for each week for each team over the season from publicly available sources. This includes injuries from the same team for each week, which is repeated data. They then calculated the mean per week per team. They had 32 teams and therefore have 32 means for a season. They then compared the mean of those means between seasons using an unpaired *t* test. First, this analysis totally ignores complications due to nonindependence in repeated data. Second, how can we understand the comparison of the means of means? Third, they compared each possible pairs of years. They ignored the multiple comparison issue. This analysis is totally inappropriate. I am not going to accept the results of this analysis, or any conclusion based on these results. This is an issue that cannot be rescued by a revision.

The authors say they have done a similar analysis in their precious paper [2]. I now doubt the findings published there too. Unfortunately, that paper was also published in *JMIR*. I recommend that editors should consider rereviewing that paper by an independent statistical reviewer.

There are less severe issues as well. For example, they presented 2 figures—one is redundant in the presence of the other, because the numbers in Figure 1 divided by the number of weeks are the numbers in Figure 2. Further, none of the numbers in any of these figures are the outcome measure they used in the statistical analysis. Therefore, the usefulness of them is limited only to describing the raw data.

Even if the analysis is correct, they have a fundamental limitation in their interpretation of the results. Their conclusions are based on the underlying assumption that the observed statistical differences were driven by training opportunities. There was no justification for that assumption. How can the authors claim none of the other possible influencing factors changed?

Conflicts of Interest

None declared.

References

1. Puga TB, Schafer J, Thiel G, et al. COVID-19 National Football League (NFL) injury analysis: follow-up study. JMIRx Med 2024;5:e45688. [doi: 10.2196/45688]

2. Puga TB, Schafer J, Agbedanu PN, Treffer K. COVID-19 return to sport: NFL injury prevalence analysis. JMIRx Med 2022 Apr;3(2):e35862. [doi: 10.2196/35862] [Medline: 35511457]

Edited by E Meinert; submitted 02.01.24; this is a non–peer-reviewed article;accepted 02.01.24; published 13.02.24.

Please cite as:
Samaranayaka A
Peer Review of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"
JMIRx Med 2024;5:e55997
URL: https://xmed.jmir.org/2024/1/e55997
doi:10.2196/55997

Peer Review of “The Role of Animal-Assisted Therapy in Enhancing Patients’ Well-Being: Systematic Study of the Qualitative and Quantitative Evidence”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/51787

Companion article: https://med.jmirx.org/2024/1/e55899

Companion article: https://med.jmirx.org/2024/1/e51787

(JMIRx Med 2024;5:e56440) doi:10.2196/56440

KEYWORDS

animal-assisted therapy; pet therapy; outcome assessment; policies; systematic study

This is a peer-review report submitted for the paper “The Role of Animal-Assisted Therapy in Enhancing Patients’ Well-Being: Systematic Study of the Qualitative and Quantitative Evidence.”

Round 1 Review

Dear Authors,

First of all, your work’s [1] topic is up-to-date and meticulously prepared. However, I still have a few questions/suggestions:

1. In the Identification section, the total number of articles obtained from each database is given. It is recommended to give separate numbers for each.

2. In the box below, the numbers are given as a total, but it may be more appropriate to give separate data for each item.

3. Can keywords be schematized in accordance with PICOS (Population, Intervention, Comparison, Outcomes, Study Design) in the literature review section?

Table …: Keywords used while browsing.

Population

Intervention

Comparison

Outcomes

Study design

4. Has the quality of evidence been evaluated? If so, how was it done? This process can be explained by creating such a subtitle.

- How did you reduce the risk of bias in studies? Were the articles evaluated and scored separately among authors? Have these scores been analyzed?
- By whom and how was the screening done? I think that the most important limitation of this study is that it was scanned by a single person.

5. In the section where general information is given for the last 16 articles, can it be added which disciplines are studied in particular? Since this subject is studied by various job groups, adding this information can enrich the data. If the mentioned situations are arranged, your article will contribute more to the literature.

6. What has been studied in previous systematic reviews? What are the original aspects of this work?

I include below some systematic review studies that may be relevant to the subject:

- Stern C, Lizarondo L, Carrier J, et al. The experiences and effectiveness of canine-assisted interventions (CAIs) on the health and well-being of older people residing in long-term care: a mixed methods systematic review protocol. PROSPERO. 2020. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020161235
- Whear R, McGill P, Orr N, et al. What are the effects of ‘robopets’ on the health and wellbeing of older people resident in care homes? A systematic review of qualitative and quantitative evidence. PROSPERO. 2017. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017081794
- Nabi N, McAloney-Kocaman K, Fleming M, Bain S. A systematic review exploring the effectiveness of animal-assisted therapy in improving the psychological well-being of incarcerated individuals. PROSPERO. 2022. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022314341

If the mentioned situations are arranged, your article will contribute more to the literature.

I wish you good luck in your work.

Conflicts of Interest

None declared.

Reference

1. Pandey RP, Himanshu, Gunjan, Mukherjee R, Chang C. The role of animal-assisted therapy in enhancing patients' well-being: systematic study of the qualitative and quantitative evidence. JMIRx Med 2024;5:e51787. [doi: 10.2196/51787]

Abbreviations

PICOS: Population, Intervention, Comparison, Outcomes, Study Design

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 18.03.24.

Please cite as:
Anonymous
Peer Review of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence"
JMIRx Med 2024;5:e56440
URL: https://xmed.jmir.org/2024/1/e56440
doi:10.2196/56440

Peer Review of “Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.08.14.23294073v1

Companion article: https://med.jmirx.org/2024/1/e57116

Companion article: https://med.jmirx.org/2024/1/e52198

(JMIRx Med 2024;5:e57701) doi:10.2196/57701

KEYWORDS

TRICARE; health care fraud; Defense Health Agency; fraud; fraudulent; insurance; coverage; beneficiary; beneficiaries; background check; background checks; demographic; security clearance; FDA; Medicaid; Medicare; provider; provider referral; military; False Claims Act; HIPAA breach; OIG-LEIE; inspector general; misconduct; insider threat; information system; zero trust; data management; Food and Drug Administration; Health Insurance Portability and Accountability Act breach; Office of Inspector General's List of Excluded Individuals and Entities

This is the peer-review report for “Insider Threats to the Military Health System: A Systematic Background Check on TRICARE West Providers.”

Round 1 Review

General Comments

In general, the manuscript [1] is informative and includes a lot of information on health care providers who participate in TRICARE insurance. The study examines those who have received some sort of exclusion, sanction, or other reprimand based on health care fraud or harm. This study is timely and has practical implications for protecting patient care, particularly for those who are in a vulnerable position such as veterans or warfighters. I hope the following comments are taken as constructive criticism and interest in the overall improvement of the study. I appreciate the opportunity to review this study.

Below is a list of important fixes that I recommend considerable time be spent on and some minor fixes. In general, I think the key limitation of the study is that it can better state the significant contribution of the study. I understand the need for such a study, but as it stands, the study can further improve by spending more time on why and how health care providers land on exclusion lists such as the Office of Inspector General’s (OIG) List of Excluded Individuals and Entities (LEIE). Indeed, the study uses several databases that exclude physicians or provide reasons why a physician no longer participates in such programs, but the author can improve their justification for the study on why this is needed.

The second key limitation of the study is the Methods and Results section. In particular, this section needs improvement with clearer detail and justification on why the author had a selection criterion (vs examining all zip codes). In addition, the Results section can improve with better organization of the findings. As it reads, the results are a bit difficult to follow with all the zip codes laid out.

Last, the study could benefit from greater discussion on the implications of the study. At the moment, it pushes for more transparency, but the author could use their data more to discuss the impact of their findings. For example, why would publishing the National Provider Identification (NPI) numbers help patients? What do patients or the author want to gain from that transparency? How can this help future patients or hold physicians more accountable? The discussion loosely taps into the implications, but the study could really tease out this argument more.

Overall, the study was easy to follow and did provide some interesting content to consider. I think the study can better serve the public and has great implications! I would like to see these implications highlighted more so that the reader can really see the contribution the study makes.

Specific Comments

Major Comments

1. Introduction: Provide an explanation of what the OIG LEIE is for the reader. It is important to inform the reader that the OIG LEIE excludes participation in the program for various reasons—not just a quality-of-care issue. For example, the OIG LEIE also can exclude physicians on a financial offense matter. This helps the reader understand the gravity of the situation, particularly when the author discusses the increased risk of mortality and hospitalization of these patients. In short, I would like to see more development and background on how individuals are included in the LEIE to increase awareness for the reader.

   a. For more information about LEIE and how physicians are placed on the list, please see the following: Burton B, Sun D, Jesilow P, Pontell HN. Two paths, one destination: a demographic portrait of physicians sanctioned by the federal government. *J Health Hum Services Adm*. 2022;45(3):142-180.

2. Methods: This section needs improvement. First, please provide more justifications and in-text citations to justify the methods used for the study. This will help strengthen the Methods section. As it reads now, there seem to be no prior studies listed that use this method (although that is not the case). Second, why did the author limit the search to the "83 most populous zip codes"? Why not include all zip codes? Does this relate to the number of people participating in TRICARE, or is this because there are simply more people living in those zip codes? Please include a justification here on why there is a population cutoff. Third, on page 8, the author writes that there were 22 states that were included, but the list only included 21 states (from my understanding). In addition, why were some of the zip codes (eg, St. Louis, Rock Island Arsenal) excluded and others included (eg, Amarillo, Lubbock, and El Paso areas only for Texas)?
3. Results: The author generally states that their findings are consistent with past research but do not include a list of articles to which they are referring. Only one article is referenced [2]. Please provide further support for that claim (in other words, please include all other studies to support the claim of consistent findings). In addition, when discussing results like on page 10, the presentation is difficult to follow with all the zip codes listed and separated by a hyphen. Please consider reorganizing this presentation or placing the list of zip codes in a footnote to ease the presentation of results.
4. Discussion: The significance of the study could be further elaborated on. At the moment, it pushes for more transparency, but the author could use their data more to discuss the impact of their findings. For example, why would publishing the NPI numbers help patients? What do patients or the author want to gain from that transparency? How can this help future patients or hold physicians more accountable? The discussion loosely taps into the implications, but the study could really tease out this argument more.

Minor Comments

1. Clean up the grammar and punctuation. For example, on page 4, the author states, "Nicholas et al performed a cross-sectional study of 8204 Medicare beneficiaries who received care from excluded providers. It revealed that patients treated by fraudsters experience a 13%-23% increased risk of mortality and 11%-30% higher risk of hospitalization (Nicholas et al, 2019)." Note, that the start of the sentence, "Nicholas et al" needs a period and a year in the citation.
2. I suggest adding a numerical list when discussing the different databases that are available for searching a physician. For example, on page 5, the author lists several different databases starting with the sentence "Multiple public databases exist to search names with respect to each of these issues, including..." Adding in a numbered list can make the information more digestible for the audience. This can also be cleaned up (ie, adding a numeric list) on page 7 when listing the different databases that the physicians were screened in.
3. Page 6, it is stated that 203 names appeared in up to 3 additional types of databases. However, what are these 3 additional types of databases? Is it referring to the earlier-mentioned databases? This is unclear.

Conflicts of Interest

None declared.

References

1. Bychkov D. Insider threats to the military health system: a systematic background check of TRICARE West providers. JMIRx Med 2024;5:e52198. [doi: 10.2196/52198]
2. Chen A, Blumenthal DM, Jena AB. Characteristics of physicians excluded from US Medicare and state public insurance programs for fraud, health crimes, or unlawful prescribing of controlled substances. JAMA Netw Open 2018 Dec 7;1(8):e185805. [doi: 10.1001/jamanetworkopen.2018.5805] [Medline: 30646294]

Abbreviations

LEIE: List of Excluded Individuals and Entities
NPI: National Provider Identification
OIG: Office of Inspector General

Edited by E Meinert; submitted 23.02.24; this is a non–peer-reviewed article;accepted 23.02.24; published 09.04.24.

Please cite as:
Anonymous
Peer Review of "Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers"
JMIRx Med 2024;5:e57701
URL: https://xmed.jmir.org/2024/1/e57701
doi:10.2196/57701

Peer Review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"

Anonymous

Related Articles:

Companion article: https://arxiv.org/abs/2304.10512

Companion article: https://med.jmirx.org/2024/1/e57838

Companion article: https://med.jmirx.org/2024/1/e48519

(JMIRx Med 2024;5:e58320) doi:10.2196/58320

KEYWORDS

opioid; substance use; substance use disorder; social media; US; opioid crisis; mental health; substance misuse; crypto; dark web; users; user perception; fentanyl; synthetic opioids; United States

This is the peer-review report for "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study."

Round 1 Review

This study [1] can be considered for publication if the researchers are able to revise it to improve the clarity of the objective, theoretical relevance, and practical value. I will suggest that there should be a segment on the objective of the study immediately after the introduction. This will help in giving the study a direction. Please, include this citation:

Obosi AC, Fatunbi AM, Oyinloye O. Peer pressure and substance use as predictors of mental health among in-school adolescents in Nigeria. *Ianna J Interdisciplinary Stud*. 2022;4(1):1 - 9.

Explain the implication of your findings to other countries, that is, give the study an international outlook.

Conflicts of Interest

None declared.

Reference

1. Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A. Detecting substance use disorder using social media data and the dark web: time- and knowledge-aware study. JMIRx Med 2024;5:e48519. [doi: 10.2196/48519]

Edited by E Meinert; submitted 12.03.24; this is a non–peer-reviewed article;accepted 12.03.24; published 01.05.24.

Please cite as:
Anonymous
Peer Review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"
JMIRx Med 2024;5:e58320
URL: https://xmed.jmir.org/2024/1/e58320
doi:10.2196/58320

Peer Review of “Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study”

Francesco Baglivo, MD

Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy

Related Articles:

Companion article: http://preprints.jmir.org/preprint/50803

Companion article: https://med.jmirx.org/2024/1/e56441

Companion article: https://med.jmirx.org/2024/1/e50803

(JMIRx Med 2024;5:e56496) doi:10.2196/56496

KEYWORDS

artificial intelligence; adoption; acceptance; opinion; perceptions; survey; expectations; physician; medical survey; qualitative study; AI

This is the peer-review report for “Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study.”

Round 1 Review

General Comments

The manuscript [1] delves into the perspectives of Brazilian physicians on the integration of artificial intelligence (AI) in medical practices through an online cross-sectional survey.

Specific Comments

Major Comments

1. The study purports to evaluate the acceptance of AI by physicians, but the specific types of AI technologies explored remain ambiguous. Are they examining generative AI, natural language processing tools, classical machine learning, or other uses of AI?
2. The phrase “Although scarcely used in real practice” comes across as too assertive. Consider a softer phrasing.
3. The methods section should provide a more comprehensive breakdown of the questionnaire’s design process. Which question types were chosen (Likert scale, yes or no, or numerical), and for what reasons?
4. When presenting results, always give raw data (numerator/denominator) along with percentages, especially after statements such as “Most of them described their AI knowledge as intermediate.”
5. There is a noticeable omission of a power analysis. How can we ascertain that the sample size sufficiently represents the broader population? The description of the target population needs elaboration.

Minor Comments

1. The statement “Artificial intelligence (AI) applied to Medicine has been a trending subject in recent years” is preferable over mentioning it as the “hottest topic.”
2. In Table 1, the age bracket should read “50-65” as the “50” seems to be missing. Several *P* values appear without context, for instance: “10. General AI Knowledge (n=164); *P*=.2565,” “11. Regularity of AI tool usage in daily life (n=164); *P*=.9792,” and “12. Familiar with medical AI solutions? (n=164); *P*=.2774.”
3. Tables 2 and 3 also contain *P* values that require explanations or clarifications.

Conflicts of Interest

None declared.

Reference

1. Giavina-Bianchi M, Amaro Jr E, Machado BS. Medical expectations of physicians on AI solutions in daily practice: cross-sectional survey study. JMIRx Med 2024;5:e50803. [doi: 10.2196/50803]

Abbreviations

AI: artificial intelligence

Edited by E Meinert; submitted 17.01.24; this is a non–peer-reviewed article;accepted 17.01.24; published 25.03.24.

Please cite as:
Baglivo F
Peer Review of "Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study"
JMIRx Med 2024;5:e56496
URL: https://xmed.jmir.org/2024/1/e56496
doi:10.2196/56496

Peer Review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"

Anonymous

Related Articles:

Companion article: https://arxiv.org/abs/2304.10512

Companion article: https://med.jmirx.org/2024/1/e57838

Companion article: https://med.jmirx.org/2024/1/e48519

(JMIRx Med 2024;5:e58321) doi:10.2196/58321

KEYWORDS

opioid; substance use; substance use disorder; social media; US; opioid crisis; mental health; substance misuse; crypto; dark web; users; user perception; fentanyl; synthetic opioids; United States

This is the peer-review report for "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study."

Round 1 Review

Comments for Authors

1. The paper [1] is well written and easy to understand. See comments below for a summary description of the paper from my perspective.
2. However, I would have liked to see insights ideally established in the medical literature and supported by the experimental context in this paper (eg, those that can substantiate the prediction results and how this type of artificial intelligence can benefit substance use disorder [SUD]–related outcomes).
3. Although a temporal pattern–aware method is implemented in this paper, which is a big positive, I would like to see an analysis over two distinctly separate time periods to establish the consistency and robustness of the proposed approach.
4. Without addressing points 2 and 3, the utility of this work is fairly limited. I would suggest a detailed discussion of points 2 and 3 in a revised version of the paper before submission.

Paper Summary

This paper presents a novel approach to SUD from social media posts crawled from various dark web sources. The pipeline is sufficiently novel and high-performing compared to the presented baselines and generally in isolation (80% plus is a good score). The authors specify the intended outcome of the study as establishing a relationship between the mention of drugs in posts versus SUD by analysis of the form of expression. The methodology, successes, and failures in detection are clearly stated and discussed.

Conflicts of Interest

None declared.

Reference

1. Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A. Detecting substance use disorder using social media data and the dark web: time- and knowledge-aware study. JMIRx Med 2024;5:e48519. [doi: 10.2196/48519]

Abbreviations

SUD: substance use disorder

Edited by E Meinert; submitted 12.03.24; this is a non–peer-reviewed article;accepted 12.03.24; published 01.05.24.

Please cite as:
Anonymous
Peer Review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"
JMIRx Med 2024;5:e58321
URL: https://xmed.jmir.org/2024/1/e58321
doi:10.2196/58321

Peer Review of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study"

Anonymous

Related Articles:

Companion article: http://preprints.jmir.org/preprint/43341

Companion article: https://med.jmirx.org/2024/1/e59637

Companion article: https://med.jmirx.org/2024/1/e43341

(JMIRx Med 2024;5:e59639) doi:10.2196/59639

KEYWORDS

COVID-19; pandemic; health care disparities; intensive care unit; ICU; right to health; quality of care; intubation; mortality; health disparity; health inequality; surveillance data; in-patient; mortality; COVID-19 patient; hospitalization; disparity; inequality; surveillance; health care system; Greece; region; Delta; Omicron; vaccination; vaccine; public health; patient load; deterioration; time

This is the peer review report for "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study."

Round 1 Review

General Comments

This paper [1] is interesting but needs improvement.

Specific Comments

Major Comments

We aimed to update this analysis to include the large "Delta" and "Omicron" waves that affected Greece during 2021 - 2022

So why did you analyze data also from 2020?

Vaccination did not affect the mortality of these already severely ill patients.

Did you check how many months before they had received the vaccine? How many of the participants were not vaccinated?

Vaccination did not show a statistically significant association with mortality, regardless of the number of doses received (Figure 2).

This needs rephrasing if all participants were vaccinated at least once.

Between 1 September 2020 and 3 April 2022...

I see that your previous study included data up to May 2021. Thus, there is a lot of overlap in the data of these two studies.

There are many errors in the use of English and typos which makes it difficult to understand. See examples below:

Mortality was significantly higher above 400 patients

This is not very clear (please rephrase).

...with an adjusted Hazard Ratio of 1.22, 95% CI: 1.09 - 1.38)

Where does the parenthesis open?

...rising progressively up to 1.48 (95% CI: 1.31 - 1.69) for 800+ patients.

This is not clear.

...we found no statistically association between mortality and...

Please correct typo.

Minor Comments

It would be good to include DOIs to all references.

Conflicts of Interest

None declared.

Reference

1. Lytras T. Health care system overstretch and in-hospital mortality of intubated patients with COVID-19 in Greece from September 2020 to April 2022: updated retrospective cohort study. JMIRx Med 2024;5:e43341. [doi: 10.2196/43341]

Edited by E Meinert; submitted 17.04.24; this is a non–peer-reviewed article;accepted 17.04.24; published 10.06.24.

Please cite as:
Anonymous
Peer Review of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study"
JMIRx Med 2024;5:e59639
URL: https://xmed.jmir.org/2024/1/e59639
doi:10.2196/59639

Peer Review of “Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/56759

Companion article: https://www.medrxiv.org/content/10.1101/2024.01.11.24301001v1

Companion article: https://med.jmirx.org/2024/1/e56759

(JMIRx Med 2024;5:e62676) doi:10.2196/62676

KEYWORDS

density; teeth; tooth; dental; dentist; dentists; dentistry; oral; tissue; enamel; dentin; Hounsfield; pathology; pathological; radiology; radiological; image; images; imaging; teeth density; Hounsfield unit; diagnostic imaging

This is the peer-review report for ”Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis.”

Round 1 Review

General Comments

The subject is interesting. The densities of dental hard tissues were determined by cone-beam computed tomography (CBCT), a technique that has been recently used for this purpose.

Minor Comments

1. The article [1] specifies the aim and is structured according to the journal’s recommendations.

2. The *Introduction* could be improved by adding some data regarding the hard structures analyzed and an appreciation of the clinical relevance of the method.

3. In the *Methods* section, the authors could compare the densities of hard tissues in the same patient by groups of teeth depending on the period of tooth bud formation.

4. The conclusions should be improved with a statement regarding the importance of the method for current practice and for its automatic use.

Conflicts of Interest

None declared.

Reference

1. Reshetnikov A, Shaikhattarova N, Mazurok M, Kasatkina N. Dental tissue density in healthy children based on radiological data: retrospective analysis. JMIRx Med 2024;5:e56759. [doi: 10.2196/56759]

Abbreviations

CBCT: cone-beam computed tomography

Edited by E Meinert; submitted 28.05.24; this is a non–peer-reviewed article;accepted 28.05.24; published 20.06.24.

Please cite as:
Anonymous
Peer Review of “Dental Tissue Density in Healthy Children Based on Radiological Data: Retrospective Analysis”
JMIRx Med 2024;5:e62676
URL: https://xmed.jmir.org/2024/1/e62676
doi:10.2196/62676

Peer Review of “Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/48213

Companion article: https://med.jmirx.org/2024/1/e60266

Companion article: https://med.jmirx.org/2024/1/e48213

(JMIRx Med 2024;5:e60429) doi:10.2196/60429

KEYWORDS

primary health care; mother and child health; community health worker; slums; digital applications; health communication.

This is a peer review for “Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study.”

Round 1 Review

General Comments

This paper [1] addresses a principal issue, especially for the developing world where the valuable lives of mothers and children can easily be prevented. However, of course, a big challenge in the proposed solution is the availability of Android devices that are also connected to the internet. This is a limitation, therefore, to be added. Another area that needs to be addressed is related to cultural acceptance and sensitivity to using technology, particularly during the prenatal stage. Also, I noted that the diagrams are not clearly developed and placed in the appropriate places. I am happy with the qualitative part but unfortunately not an authority on quantitative; thus, this part should be vetted by a quantitative expert.

Specific Comments

Major Comments

1. This is an excellent topic requiring continuous literature development.

2. The method used is mixed, whereas I would have preferred the total qualitative inquiry considering the set aims and objectives.

3. The authors need to pay attention to the sociocultural realities of the context; therefore, either address them or acknowledge them as limitations.

Minor Comments

4. The diagrams need to be appropriately designed and placed in the paper.

Conflicts of Interest

None declared.

Reference

1. Haq ZU, Naeem A, Zaeem D, Sohail M, Pervaiz NUA. Development of a digital platform to promote mother and child health in underserved areas of a lower-middle-income country: mixed methods formative study. JMIRx Med 2024;5:e48213. [doi: 10.2196/48213]

Edited by E Meinert; submitted 10.05.24; this is a non–peer-reviewed article;accepted 10.05.24; published 31.07.24.

Please cite as:
Anonymous
Peer Review of "Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study"
JMIRx Med 2024;5:e60429
URL: https://xmed.jmir.org/2024/1/e60429
doi:10.2196/60429

Authors' Response to Peer Reviews of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"

Izzy Kohler[1], DPT; Michael V Perrotta[1], BSc; Tiago Ferreira[1], MS; David M Eagleman[1,2], PhD

[1]

[2]

Corresponding Author:
David M Eagleman, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/49969

Companion article: https://www.medrxiv.org/content/10.1101/2023.06.01.23290351v1

Companion article: https://med.jmirx.org/2024/1/e55554

Companion article: https://med.jmirx.org/2024/1/e55727

Companion article: https://med.jmirx.org/2024/1/e55728

Companion article: https://med.jmirx.org/2024/1/e49969

(JMIRx Med 2024;5:e55510) doi:10.2196/55510

KEYWORDS

audiology; hearing; high-frequency; wristband; develop; development; wearable; wearables; machine learning; phoneme; phonemes; hear; vibrotactile; vibration; vibrations; sound; sounds; hearing loss; loud noise; loud noises; noise pollution; hearing aids; hearing aid

This is the authors' response to peer-review reports for "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation".

Round 1 Review

We thank the reviewers for their very helpful feedback. Following their suggestions, we have clarified the language throughout and added several new figures and tables. Collectively, this has strengthened the manuscript and should address all concerns. Detailed responses below.

Reviewer F [1]

The authors report on an interesting study [2] in which they use a wearable device to sense high-frequency sounds. I have some specific comments below. To summarize, some essential elements are missing from the manuscript, and the manuscript needs significant editorial attention (errors, academic writing style, figures).

Introduction: I would suggest using primary references for the number of people with hearing loss (rather than Olusanya et al [3]) and for the burden of hearing loss (rather than Michels et al [4]). Regarding the risk of high-frequency hearing loss, have the authors overlooked the fact that this is commonly seen in most older adults (ie, what is attributed to aging)? This is mentioned in the second paragraph. The authors are mixing up noise-related hearing loss and age-related hearing loss (presbycusis) in the manuscript.

Response: Thank you for all your comments. The Introduction was reworded for clarity. Additional references have also been added throughout the Introduction section. All is detailed below.

Last sentence on the first page should not finish with a colon.

Response: This has been replaced with a period.

I do not think that Hickson et al [5] is a primary reference for limitations of hearing aids (HAs) and cochlear implants. "The auditory cortex is activated by vibrotactile information in individuals who are hearing impaired and deaf." This implies that the auditory cortex is only activated this way.

Response: This sentence has been revised to "The auditory cortex is primarily dedicated to the processing of sound, but can also be activated by vibrotactile information in individuals who are hearing impaired and deaf [6,7]."

Middle paragraph: phonemes are extracted. How this is done should be provided here, not later in the manuscript.

Response: All information regarding phoneme choice is presented together under the Algorithm section.

Is the designation of the particular transducer important? In other words, is a larger temporal difference between the two most similar phonemes important?

Response: This question was not tested in this research study; each phoneme was simply assigned to a different actuator. Our previous research [8] demonstrates that participants can learn to distinguish the spatial differences.

"...unconsciously integrated..." Why is the word "unconsciously" needed?

Response: With practice, the integration of vibrations with sound becomes automatic, not requiring constant awareness of which actuator is vibrating and the phoneme assigned to it.

"The user is then able to understand..." Isn't this yet to be shown, or is evidence provided in the next paragraph? If so, this needs to be made clearer.

Response: We have added citations to previous published work to make this clear.

Interestingly, the microphone is placed on the wrist, a part of the body that can often be situated away from the direct line of communication between two people (eg, under a table). Were the users trained to keep their wrists up?

Response: Users were not trained to hold their wrist in a specific position; this was unnecessary because the microphone is omnidirectional.

". (Phatak et al., 2009) asked" and "s. (Sher & Owens, 1974) presented...": references not properly incorporated into the sentences

Response: These have been corrected, thank you.

"...listening to an audiobook, podcast..." These are often streamed to personal headsets/earphones. Were any instructions provided in terms of volume, closeness to speakers, etc?

Response: Participants were asked not to use headsets or earphones so that the microphone on the wristband would collect the sound. No further directions were given for volume or closeness to speakers. This has been clarified in the manuscript.

Participants: Normally, information about participants is provided before most of the other information in a Methodology section, particularly before, for example, tasks.

Response: The order of presentation within the Methods section has now been changed to place participants first.

Abbreviated Profile of Hearing Aid Benefit (APHAB): I am not sure that I agree with the rationale that questions on aversiveness are not relevant. Cox and Alexander [9] write "Aversiveness of Sounds, quantifies negative reactions to environmental sounds," and "The APHAB is a potentially valuable clinical instrument. It can be useful for quantifying the disability associated with a hearing loss and the reduction of disability that is achieved with a hearing aid." That is, it is designed to be used before an intervention (and has been used a lot for non-HA interventions as well, eg, implants).

Response: Thank you for the comment. We removed aversiveness questions from the APHAB because the haptic wristband does not alter or distort sound (as HAs can), and therefore, these questions did not directly apply to a wristband (eg, changing one's tolerance for different types of sounds). We have clarified this in the manuscript.

How was the APHAB administered?

Response: The test was administered through an online questionnaire that captured the data onto a datasheet for analysis. This information has been added to the manuscript under "Abbreviated Profile of Hearing Aid Benefit (APHAB)"

Please pay attention to tense when writing. In most cases, past tense should be used.

Response: Done, thank you.

How many male and female participants were in the study?

Response: 10 males, 5 females, and 1 nonbinary. This information has been added to the Participant section.

dB should be dB hearing loss.

Response: This has been corrected.

Using an audiogram from any mobile-based device means little guarantee of accuracy.

Response: We have now clarified in the manuscript that smartphone hearing apps (eg, Mimi; which we used) has been found to be comparable to in-clinic testing (eg [10]).

What was the rationale for the specifications for the audiogram?

Response: This was simply a general inclusion criterion to make certain we were capturing garden-variety presbycusis.

"ear's" should be "ears."

Response: corrected

Any reason why 16 people were recruited?

Response: As a general rule, we consider 10 subjects a minimum number for a good psychometric study. In this case, we recruited 19, and 3 dropped out. Our retrospective power analysis shows that 16 participants were well sufficient given the outcome magnitude.

What is "(11.6)," SD? And "(13)" and "(9)"?

Response: Yes, we have now clarified this in the manuscript.

Figure 3: I suggest not including the values on the plot. Furthermore, "Error boundary represents standard error of the mean." The reader has to interpret the "error boundary" as the gray area.

Response: We would prefer to keep the values in the plot, as more information is better. However, we have clarified the definition of error boundary in the figure caption.

"...to drop at a slower, more steady pace for the remaining five weeks of the study." Writing could be tightened up a bit, and there is a rise in scores at 3 weeks. If the response is that there is not a significant increase, then it would be good to report at what point the difference is not significant.

Response: This was reworded to "The average aided APHAB score continued to trend down for the remaining 5 weeks of the study." The wording was chosen so as not to imply that continued improvement stopped after a week.

Regression analysis: This is OK, but the use of a paired sample t test could have been taken for both analyses.

Response: There may be a misunderstanding here. The regression analysis in Figure 6 of the final paper simply characterizes the relationship between baseline score and final outcome. A *t* test would not be possible here.

On the other hand, a multinomial regression analysis could have considered the influence of age, HA user or not, or baseline APHAB scores on final APHAB scores.

Response: Thank you for the suggestion. Unfortunately, our sample size is not sufficiently large enough to yield good signals from a multinomial regression, especially as some of the suggested categories ("HA user or not") are binary. We will keep this in mind for future studies as our sample size grows.

I see that there were approximately equal numbers of HA and non-HA users. Was this by accident or design? It is not mentioned in the Recruitment section.

Response: The approximately even numbers was hoped for but fortuitous.

Do not start sentences with "Also."

Response: We have replaced with "additionally."

Figure 6: It is not clear which score is being reported. At 6 weeks? I suspect it means the difference between the baseline and final scores. If so, this needs to be made clear in the caption.

Response: The caption has been corrected for clarification; the graph represented 6 weeks.

It appears that there was no attempt to record the listening environments of the users nor how often they used their devices.

Response: We have added the following segment to the Results section:

"Time wearing the wristband and time exposed to speech was verified through collection of data from backend logging that records when the wristband is turned on or off and when a phoneme is detected. As seen in Figure 5 participants wore the wristband for and average of 12.9 (SD=8.1) hours per day and were exposed to speech for an average of 6.7 (SD=3.3) hours per day."

"One potential hypothesis" should be "One potential explanation."

Response: This has been corrected.

"Participants without hearing aids benefitted the most from vibrotactile sensory substitution..." True—in fact, those with HAs did not get significant benefits.

Response: Thank you. This is now described in detail in the Results section.

It is always good to devote a bit of space to the limitations of the study. This is missing in this manuscript.

Response: Thank you. A limitations paragraph has now been added.

"Future studies will focus on quantifying the maximum benefits possible and how long improvements continue before a plateau is reached." This is not a conclusion of the study.

Response: Thank you. This line was removed from the Conclusion.

Perhaps this is mentioned elsewhere, but the device is given a name; it would be good to know about the association between the authors and the manufacturer of the device.

Response: The authors are associated both with Stanford University and the company Neosensory, which makes this device. This information is in the paper.

Anonymous [11]

General Comments

This paper highlights the utility and perceived communication benefits of the Clarity vibrotactile band for users with high-frequency hearing loss. Overall, this is a well-designed study that demonstrates the effectiveness of this assistive listening device that provides benefits for listeners with high-frequency hearing loss in complex listening situations as measured by the APHAB. Additionally, this study provides subjective evidence that both HA users and non-HA users experience benefit from the Clarity device. Specifically, the non-HA users report more benefits across different listening conditions (background noise [BN] and reverberation) than HA users.

Major Comments

1. *Consider referencing Glick and Sharma [12] in your Introduction as it relates to the cross-modal plasticity associated with age-related hearing loss (presbycusis).*

Response: This reference has been cited. Thank you for making that recommendation.

2. *In the Methods section, consider starting with a clear description of the participants. Who are they, how many, how many were HA users versus non-HA users, age, etc. While the majority of this information is embedded later in the article, it is not readily accessible.*

Response: A demographic chart was added to the manuscript that outlines the important demographic characteristics of all of the participants.

3. *In the Methods section, consider creating a subheading or table for the audiometric data of the participants and including additional information like a description of their audiometric*

data (type, degree, configuration), pure tone average (500, 1000, and 2000 Hz), symmetry of the hearing loss, how many were considered to be within normal limits up to 2000 Hz versus having hearing loss at lower frequencies (≤2000 Hz). This could have a significant impact on speech understanding difficulties, especially in complex listening environments.

Response: A chart has been added to the supplementary materials (see Table 1 in final paper); it includes all audiometric data for the participants.

4. *For the audiometric data, how many participants provided their test results from a doctor of audiology or hearing health care professional? How many provided results from the mobile app? Is it possible to confirm that all participants had sensorineural hearing loss and not mixed or conductive hearing loss?*

Response: A total of 7 participants provided 2 audiograms from the online assessments (a Mimi hearing assessment), and 9 provided audiograms from an audiologist. The type of hearing loss was not confirmed; this has been added to the Limitation section of our Conclusion.

5. *In the Device subsection, consider adding additional information regarding the microphone characteristics. Additionally, define "GRMS."*

Response: A table was added to the supplementary materials (Multimedia Appendix 1 in final paper).

6. *In the Algorithm subsection, you mention the sham algorithm and the /f/ motor. In the sham condition, which motor represents the /f/ phoneme, and which additional phonemes are used in the sham condition?*

7. *Additionally, the sham condition is never mentioned in the Results or Discussion. Consider adding this information to the manuscript, or if you choose not to, consider not introducing the sham algorithm.*

Response: Oops, that sentence was mistakenly included from a previous internal study. We have fixed this now, removing the description of the sham algorithm. For clarity, in this experiment, a sham was not used.

8. *In Figure 3, consider changing the y-axis to "APHAB Score (%)" and refer to the APHAB benefit scores as scores or percentages instead of points in the text.*

Response: The standard method of interpreting the APHAB is to look at unaided (baseline), aided (final), and benefit scores (unaided – aided). Please see the following paper for more details:

Cox RM. Administration and application of the APHAB. *Hearing J.* Apr 1997;50(4): 32. [doi: 10.1097/00025572-199704000-00002]

9. *For the simple linear regression, consider adding a statement that indicates what this means or its importance.*

Response: This was reworded for clarification: "Simple linear regression analysis was used to test if a participant's baseline APHAB score explains their benefit APHAB score after 6 weeks, indicating that those with greater subjective difficulty understanding speech may stand to benefit the most from the haptic assistance of the wristband."

10. *In Figure 5, consider adding bars for weeks 0 and 1 to help readers visualize the results in the text.*

Response: Thank you for this suggestion; we have added this as Figure 7 in the final paper.

11. *Consider creating a line graph that highlights the greater decrease in APHAB scores from baseline to week 6 for those without HAs than those with HAs (as discussed in the Results).*

Response: Thank you for this suggestion. The graph we added (Figure 7 in the final paper) highlights the difference as per your request.

12. *In Figure 6, this figure represents benefit scores from baseline (wk 0) to week 6, correct? Consider clarifying the figure text and removing the information regarding the subgroups.*

Response: We further clarified the figure in the caption. We prefer to keep the subgroups represented in the caption to illustrate what is further described in the text.

13. *In the Discussion and Conclusion sections, I do not think it is accurate to say that the Clarity device "improved their understanding of speech communication" because that was not what was measured. The APHAB is a subjective measure, which to me means that all the benefits users received from using the Clarity are perceived benefits and are not measurable improvements in understanding. To claim speech understanding improvements, I feel you would need to document that through an objective speech understanding measure such as the word recognition score in quiet, word recognition score in noise, Quick Speech in Noise, etc.*

Response: The Discussion and Conclusion were reworded to clarify the subjectivity of the APHAB and what the results indicate. For example, in the Discussion, we have rephrased our sentence to say "Here, we demonstrated that individuals with high frequency hearing loss are able to improve their subjective understanding of speech communication using vibrational representations of high frequency speech sounds on the wrist."

14. *In the Discussion section, you refer to the group with a higher APHAB score experiencing a greater improvement. Is this the group that uses HAs, or is this a different subgroup? It would be interesting to know how many in this group had hearing loss between 250-2000 Hz.*

Response: This was referring to the subgroup that started the study with a higher baseline score. This clarification has been added to the sentence.

15. *In the Discussion section, you report subgroup data for BN, reverberation, and ease of communication (EOC) that is not documented or reported in the Results section or any figures/tables. Consider adding this.*

Response: The scores referred to in the Discussion are all reported in the Results section. Figure 8 in the final paper is the accompanying graph.

16. *In the Conclusion section, you mention that "results also demonstrate that individuals who had the greatest amount of difficulty understanding speech prior to." Is this the without HA subgroup or a different subgroup? A few times throughout the article, these labels appear to be used interchangeably. While this may be accurate for your data set, I would caution that these terms/labels are not mutually exclusive.*

Response: Those who had the greatest amount of difficulty understanding speech prior to starting the trial refers to those who started the study with the highest APHAB baseline score. The line in question has been reworded to:

"Finally, our results also demonstrated that those who started the study with a higher APHAB score (greater hearing disability) experienced the greatest amount of benefit from vibrotactile feedback."

Minor Comments

1. *In the Introduction, the authors mention that HA and cochlear implant users commonly report disappointment with understanding speech and reference Hickson et al [5]. While this could be true, the majority of users' complaints are specifically related to difficulties understanding speech in complex or noisy listening environments, not just in quiet as is implied.*

Response: This sentence has been changed to "One of the most commonly reported disappointments among users of HAs and CIs is that they still cannot understand speech, especially in complex environments."

2. *How much were participants compensated for their participation?*

Response: Participants were given a US $100 gift card for their participation. This is now clarified in the manuscript.

3. *In Figure 2, I assume your scale for the y-axis is dB of hearing loss? Consider clarifying which dB scale was used.*

Response: Thank you, this has been corrected.

4. *In the Paradigms subsection, does the Clarity device have any data logging features that can objectively record how often or how long the participant is using the device or in what listening conditions the user is in with the device (eg, quiet rooms, noisy restaurants, or reverberant auditoriums)?*

Response: A usage graph (Figure 5 in the final paper) has been added.

5. *In the APHAB subsection, consider rewording for clarity: "modified version of the Abbreviated Profile of Hearing Aid Benefit (APHAB) which did not include six questions related to the aversiveness subscale (Cox, 1997)."*

Response: We have changed the wording, thank you.

6. *In the Results section, consider rewording for clarity: "...they ended the study at a lower level of disability than those with hearing aids."*

Response: We have reworded the sentence, thank you.

7. *The implication of microphone location briefly mentioned in the Discussion is very important in my opinion. Microphone location is a significant issue even for ear-level HAs. I can only imagine the microphone placement significantly impacts the benefit and utility of the Clarity.*

Response: This was added as a limitation of the study.

8. *In the Conclusion section, consider rewording for clarity: "We found that while both hearing aid and non-hearing aid users with high frequency hearing loss reported benefited, vibrotactile feedback appears to be more beneficial for non-hearing aid users."*

Response: Done, thank you for the suggestion.

9. *The manuscript does not include an ethical approval statement or a limitations section.*

Response: Ethical Approval section has been added.

"The study protocol was approved by Solutions IRB, an independent institutional review board accredited by the Association for the Accreditation of Human Research Protection Programs, Inc. All subjects gave written informed consent in accordance with the Declaration of Helsinki."

Limitations section has been added:

"There are limitations of this study. First, the small sample size prevents extrapolation of the results to larger populations; this will be addressed in future studies. We were also limited in our ability to collect speech comprehension data in a noise-controlled environment with standardized volume controls – this is because the testing was done in participant homes instead of a laboratory. As a result, this study depended on self-report data (APHAB) which always has the potential of being influenced by a placebo effect. Another limitation is that some participant audiograms were assessed via phone applications rather than an audiologist's office; however, it should be noted that these appear to yield roughly equivalent results [13]. We also note that the specific type of hearing loss was also not controlled for beyond meeting the audiogram requirements. One final thing to note is that participants could move their hand (and hence their wristband), meaning that the microphone placement was not standardized in a single position. We do not consider this a limitation of the study, as the study is meant to test whether a vibrotactile wristband can be used to detect sound. The positive results reported here suggest that the mobility of the microphone does not present a problem."

Anonymous [14]

Overall

This study reports on an interesting device with intriguing clinical implications for people with hearing loss.

Innovative, and worthy of reporting on this technology, which could inspire other researchers

But there are some issues that I feel require revisions:

- *Conflation of self-reported and objective benefit in the write-up*

- *Lack of reporting the range and dispersion of the data—paper focuses on group means and gives very little ability to draw any inferences about individual participant variability*
- *Lack of objective data about performance of the algorithm and participant performance for speech understanding*

Response: Objective data for algorithm performance has been added.

- *No data presented for the final questionnaire presented in the Methods*
- *Presentation and discussion of results switches back and forth between benefit scores and raw scores in a way that is unclear and makes the paper difficult to follow and interpret at times*
- *Some conclusions are presented without statistical results to support them*
- *Some conclusions are stated too strongly given the sample size and study design*
- *Lack of a limitations section to help reader contextualize the results*

Abstract

- *"...improve their understanding of verbal communication.": Please indicate that this is a self-reported or self-perceived understanding of verbal communication. I think it is important to distinguish the results from objective speech recognition testing (acknowledging that self-reported benefit is very important).*

Response: We have modified the wording to emphasize the subjective nature of the APHAB.

- *"...greatest amount of benefit...": Please indicate that it is a self-reported or self-perceived benefit.*

Response: We have modified the wording to emphasize the subjective nature of the APHAB.

Introduction

Page 2: Authors indicate that auditory and vibrotactile information can be unconsciously and naturally integrated in the brain. It would be helpful if the authors could give some description/details of how the integration is hypothesized to occur—how long it takes and what neural/cognitive mechanisms might support it. Even if this is just a hypothesis, it would provide helpful context.

Page 3: "...can help individuals with high frequency hearing loss better understand speech communication..." Please indicate that it can help them self-report or self-perceive better speech communication.

Response: Reworded to "In this study, we demonstrate that a simple wearable sensory substitution device that transforms speech sounds into haptic vibrations on the wrist can help individuals with high frequency hearing loss to feel more confident in their ability to understand speech communication throughout their normal daily routine."

Page 3: The last sentence is too strong. A device can help improve self-reported speech communication without translating to the types of benefits the authors describe in these various situations/environments. It could say something like "the evidence demonstrate the promise of this technology, which if further developed and refined holds promise for..."—something like this. I think it is OK to indicate that these kinds of benefits are possible in the future, but they are not directly supported by the results of this small study. Lots more work is needed.

Response: Reworded to "With further development and refinement, this technology has the potential to improve the quality and productivity of their daily interactions, enable them to enjoy audio based entertainment such as movies and podcasts, help them understand conversations in complicated acoustic environments, and fill the residual gaps of impairment left by their hearing aids."

Methods

Not much detail about the machine learning algorithm is provided. More detail about how it filters BN and identifies phonemes would be helpful. How was the algorithm trained? Assuming it was trained on speech, what regional accents were used?

Response: We have added the following details to the Methods section:

"The phoneme detection algorithm was trained using the elastic compute cloud on Amazon Web Services (AWS). The training data consisted of a combination of pure LibriSpeech and Librispeech re-recorded through the onboard microphone on the wristband. Librispeech is a corpus of approximately 1000 hours of English speech with standard American accents sampled at 16 kHz that has been shown to produce excellent performance in speech recognition models trained with it [15]. To produce a corpus of English read speech suitable for training speech recognition systems, Librispeech aligns and segments audiobook read speech with the corresponding book text automatically and then filters out portions with noisy transcripts. The purpose of using re-recorded data was to tune the algorithm's parameters to speech sounds representative of those it would encounter from the wristband's microphone."

Related to the above, it is not clear how the sham algorithm was used in developing the algorithm. Additional detail/description would be helpful.

Response: There was no sham algorithm in this study. That sentence was inserted mistakenly and has been removed.

Page 4: The authors mention that the algorithm performed poorly for some consonants that people with hearing loss have trouble hearing. It is not clear what level of performance constitutes poor performance and what level constitutes good performance for the phonemes that were selected for the algorithm. More context here would help the reader to understand the results. Understanding the algorithm's accuracy is important for contextualizing the users' results. It would be reasonable to suspect that the users' results should be closely linked to the algorithm's accuracy.

Response: The specifics of the machine learning algorithm's performance have been included in Table 2 in the final paper.

Tasks

I am a bit confused as to why objective speech recognition testing was not completed. The self-reported benefit is absolutely important, but based on the Introduction, the reader is interested in knowing how objective speech recognition improved with the wristband for the selected consonants. If these data are available, it would be helpful to add them. If not, it would be helpful if the authors could explain—somewhere in the manuscript—why this testing was not completed/reported.

Response: This information was added to our Limitations section: "We were also limited in our ability to collect speech comprehension data in a noise-controlled environment with standardized volume controls. This is because the testing was done in participant homes instead of a laboratory. As a result, this study depended on self-report data (APHAB), which always has the potential of being influenced by a placebo effect."

Final questionnaire: It does not seem like the results of the final questionnaire are reported in this manuscript. Given that the APHAB is the only reported outcome measure, it would be helpful to add these results as well, as they represent something more holistic than the weekly APHAB results.

Response: Three of our participants requested to continue use of the wristband after the study ended, and hence, they did not fill out the final questionnaire. Of those who did, some had criticisms ("I'm really unsure if the Clarify band was helpful or not") and some had praise ("It was very beneficial. Thank you"); however, the comments were too few to be statistically meaningful. This information has been added to the Results section.

Paradigm

Does the wristband provide any data logging to indicate how many hours per day the devices were worn? If not, this is not a major flaw but should be mentioned as a limitation because it seems like wear time could directly affect benefit.

Response: We have added the following segment to the Results section:

"Time wearing the wristband and time exposed to speech was verified through collection of data from backend logging that records when the wristband is turned on or off and when a phoneme is detected. As seen in Figure 5, participants wore the wristband for and average of 12.9 (SD=8.1) hours per day and were exposed to speech for an average of 6.7 (SD=3.3) hours per day."

APHAB

It might be helpful to the reader to clarify that higher raw APHAB scores indicate worse performance and lower scores indicate better performance but higher benefit scores represent more benefit or better outcomes.

Response: We added this line to the APHAB section: "Lower raw APHAB scores indicate lower levels of disability associated with hearing loss. Higher benefit scores indicate more perceived benefits from intervention."

Participants

I would suggest adding the number who did and did not use HAs in this section. Any additional information regarding participants—gender, education, etc—would be helpful if it is available to report. Otherwise, I suggest adding that a limitation of the paper is the limited demographic information of the participants (combined with a small n), which makes it hard to determine if any participant-level characteristics might influence the benefit of the wristband.

Response: We have added a demographic table to the paper indicating demographic information for each participant: age, gender, HA use, years with hearing loss, and hearing loss profile (Table 1 in the final paper).

The authors mention that if a clinical audiogram was unavailable, participants completed an audiogram via a mobile app. Then the authors provide an example, Mimi. Did everyone who used a mobile app use the Mimi app, or did some use other apps?

Response: The two tests used were Mimi [16] and the Hearing Test & Ear Age Test [17]. Participants who did not have an audiogram from an audiologist were required to provide audiograms from both apps. This was clarified under the Participants section.

Relatedly, it would be helpful to report how many participants had clinical audiograms and how many used an app to provide context for the audiometric results.

Response: This line was added under the Participants section: "Nine participants provided audiograms from an audiologist and seven provided audiograms from the 2 mobile apps."

Results

One critique is that I did not feel like I got a very good handle on the descriptive statistics before the authors started showing group means (with SE of the mean) and comparisons (both over time and between subgroups of participants). I felt that the results emphasized group means (with SE of the mean), but I did not get a good sense of the range and dispersion of the data. In the Discussion, the authors start discussing the numbers of participants who started or ended at a specific APHAB overall score range, but I did not feel like I had the information in the paper to help me contextualize that discussion (because the results, as presented, do not give a very clear view of how individual participants may have performed).

Response: Thank you for this. We have now added tables in the supplementary material (Multimedia Appendices 2 and 3 in the final paper).

To address the above point, I would strongly suggest adding a descriptive results table first that gives the means, maximums and minimums, and SDs for the overall APHAB scores and, possibly, APHAB benefit scores. It would be nice to see these values for the full participant group, as well as for the subgroups of participants with and without HAs (including the n in each group). It would be helpful to see the same data for the subscale scores (EOC, BN, reverberation) if it fits in the table, but I think the overall APHAB scores would be sufficient if space is an

issue. Another consideration is that with only 16 participants, you could show the individual-level data for each participant, who would each be a row, and then give the group data in a different row. I defer to the authors on their preferences but would simply suggest that some revisions be made to give the reader a better grasp of the descriptive results.

Response: Done—see above.

In the section where subscale analyses are given, the write-up describes comparisons of subscale benefit scores between the different subgroups (with and without HAs) as well as comparisons, within a subgroup, of benefit scores to the baseline score. Throughout this section, it is hard to track which P values go with which comparisons. It is hard to read and interpret. Additionally, the information is presented slightly differently for each subscale, which makes it even harder to follow. Clearer written descriptions of each comparison being tested—then followed by the statistical numbers—would be beneficial for the reader. Additionally, using a parallel results presentation for each subscale would be helpful.

Response: Thank you for the suggestions. This section was reworded with better consistency and clarity.

"Subscale analyses were performed for ease of communication (EOC), background noise (BN), and reverberation (RV) (Figure 8 and Supplemental Table 2). These subscales are reflective of speech communication under ideal conditions, in noisy environments, and in reverberant environments. The average benefit score for EOC was 15.44 (SD=13.88 , *n*=16, *P*<.001 , two-tailed dependent t-test). Those who wore hearing aids and those who did not wear hearing aids had similar EOC benefit scores (t(14)=2.18, *P*=.6, two-tailed independent t-test). The average EOC benefit for those with hearing aids was 13.57 (SD=15.71, *n*=9, *P*=.03, two-tailed dependent t-test) and the average EOC benefit for those without hearing aids was 17.83 (SD=11.85, *n*=7, *P*=.01, two-tailed dependent t-test). The average benefit score for BN was 10.88 (SD=17.54, *n*=16, *P*=.03, two-tailed dependent t-test). The average BN benefit for those without hearing aids was 16.99 points higher than those hearing aids (t(14)=2.14, *P*=.05, two-tailed independent t-test). The average BN benefit for those with HA was 3.44 (SD=17.5 , *n*=9, *P*=.54, two-tailed dependent t-test) and the average BN benefit for those without hearing aids was 20.43 (SD=15.1 , *n*=7, *P*=.01, two-tailed dependent t-test). The average benefit score for RV was 10.84 (SD=16.95, *n*=16, *P*=.02, two-tailed dependent t-test). The average RV benefit score for those without hearing aids was 11.12 points higher than those with hearing aids (t(14)=2.14, *P*=.20, two-tailed independent t-test). The average RV benefit for those without hearing aids was 17.10 (SD=16.0 , *n*=7, *P*=.03, two-tailed dependent t-test) and the average RV benefit for those with hearing aids was 5.98 (SD=17.0 , n-9, *P*=.32, two-tailed dependent t-test)."

Discussion

"...individuals with high frequency hearing loss are able to improve their understanding of speech communication...": I would like to see it be specified that this is an improvement in self-reported or self-perceived understanding of speech communication. Previous HA research shows there can be a placebo effect associated with the perception that one is wearing advanced technology [18].

Response: Language has been changed throughout the document

"...participants were able to improve their ability to understand conversations during daily interactions.": Same comment as above. Please indicate this is a self-reported or self-perceived ability to understand conversations.

Response: Language has been changed throughout the document to "Here, we demonstrated that individuals with high frequency hearing loss are able to improve their subjective understanding of speech communication using vibrational representations of high frequency speech sounds on the wrist."

"We further found that participants who started the study with a higher APHAB score experienced a greater improvement in their ability to understand speech by the end of the six week trial.": As mentioned earlier in this review, this result is hard to interpret without any sense for the individual variability in the data. The results are presented as group means without clear maximums and minimums or SDs. Providing this information in the Results would help give context to this claim in the Discussion.

Response: As above, we added two tables with this information to the supplementary material (Multimedia Appendices 2 and 3 in the final paper).

"Out of 16 participants, 14 ended the study with an APHAB score of 40 or below...": This is again difficult to interpret without any sense for the individual-level data. At timepoint zero, the group mean is right around 40. It is not clear if ending the study at 40 or below indicates benefit or is just a reflection of peoples' starting points. The discussion should be framed in terms of the amount of benefit people reported.

Response: A graph (Figure 4 in the final paper) was added to the Results section.

"Five participants started the study with an unaided APHAB of 50 points or higher...": Again, a better sense of the individual-level data and dispersion would help give context for this. The Results are focused on means and then the Discussion brings up individual data, and it is hard to interpret the two together.

Response: See the supplementary tables (Multimedia Appendices 2 and 3 in the final paper).

A small point but should this be <30 not >30 as written in the text?

Response: No, the >30 benefit score is correct.

"One potential hypothesis...": It seems like this could also be due to having more room to improve their everyday speech understanding. I think it is important to acknowledge possible noncortical factors that could explain this finding (though I think it is fine to also leave the possibility that it reflects cortical characteristics).

Response: This was reworded to "One potential explanation for why participants who started the trial with greater difficulty understanding speech experience greater improvement is that

more of their auditory cortex is available for the interpretation of tactile sound representation (Auer et al ., 2007). It is also possible that participants who started the study with a lower APHAB score had more room for improvement. This could be an interesting topic for future research."

"Participants without hearing aids benefitted the most...": I think given the lack of statistical significance in the comparison of the group means, this needs to be toned down a bit. Perhaps something like "Participants without hearing aids demonstrated a trend toward higher self-reported benefit, though this did not reach statistical significance." I know the authors reference the Cox 10-point criterion, but I am not sure that can be accurately applied to these data when the statistical test says the group means themselves are not statistically different (maybe related to the small sample size and variance in the data). Again, I would also like the benefit to be specified as self-reported or self-perceived.

Response: This has been changed to "Participants without hearing aids demonstrated a trend toward higher self-reported benefit from vibrotactile sensory substitution for speech understanding, though this did not reach statistical significance."

"Given that this group started the study with a higher APHAB score...": I did not find where there is a statistical test to justify this claim. This should be justified with a t test. Otherwise, I think it would be OK to specify that the t test did not show a statistical difference, but this group is trending toward having a higher baseline APHAB score.

Response: This was reworded to "Given that this group started the study trending toward a higher APHAB score (above), we presume the difference is because the hearing aid group already gains benefit from their technology and therefore has less room for improvement."

"In this study, we demonstrated the addition of vibrotactile feedback in the presence of background noise enabled individuals who did not wear hearing aids to hear speech communication better...": Again, would like to see it noted that this is a self-reported or self-perceived benefit.

Response: This was reworded to "In this study, we demonstrated the addition of vibrotactile feedback in the presence of background noise enabled individuals who did not wear hearing aids to hear speech communication better based on their subjective experience."

The authors present the final average BN scores (eg, 28.95 and 40.04), but the section above seems to be focused on benefit scores. This reflects my earlier comment about providing more descriptive data upfront. It is hard to track how the authors switch between baseline and benefit scores, and without a descriptive table to refer to, it is difficult to contextualize some of the Discussion.

Response: The benefit score is the baseline score minus the final score. The supplementary table (Multimedia Appendix 3 in the final paper) should help to clarify this.

Related to the above, these scores are presented as being different but are they statistically different?

Response: Indeed, they were statistically significant. This is clarified in the Results section:

"The average benefit score above baseline for BN was 10.88 (SD=17.54, n=16, P=.03, two-tailed dependent t-test), with a 16.99 point difference in BN benefit between those who wore and did not wear hearing aids (no hearing aids 20.43 benefit, hearing aids 3.44 benefit, t(14)=2.14, P=.05, two-tailed independent t-test)."

"...suggesting that those who use hearing aids may benefit from using vibrotactile feedback during conversations in background noise instead of using their hearing aids.": I think this is much too strong of a conclusion for the data, study design, and sample size. This needs to be significantly toned down—as written, I think this is a reckless conclusion based on the limitations of the data. I could be OK with presenting this as an interesting finding worth future research to determine if the above could potentially be true. However, it would need to be framed by saying that this sort of clinical recommendation would require much larger, more rigorous studies with blinding of participants and researchers.

Response: This line was added for clarification: "While our data do not offer conclusive evidence of this due to several limitations, it does offer an area worth further exploration in larger studies."

"Similar to our findings in background noise, we also found...": From what I can see in the subscale results discussion, the difference between the group with and without HAs did not reach statistical significance. If this is true, it seems to be going too far to say that the wristband helped people without HAs the most. Here, I think it is OK to note that the results are trending in this direction as long as it is acknowledged that the results did not reach statistical significance.

Response: This has been rephrased: "Here we found that the addition of vibrotactile haptic vibration to the wrist in reverberant environments tended to help the participants without hearing aids more than those with hearing aids, though the difference did not reach statistical significance."

"At the end of the trial, the group of participants who did not wear hearing aids showed an average reverberation score that was less than the average for the group who were regular hearing aid users.": Was this tested statistically? From what I can tell, it looks like only the benefit scores are presented in the Results—not the raw scores. If the Discussion brings up the raw score (not the benefit), this should be presented in the Results section. Again, statistical results are needed to draw conclusions regarding the comparison of means.

Response: This sentence was consolidated with the prior sentence (above).

"It is possible that individuals who use hearing aids may find haptic vibrations to be more helpful in reverberant environments...": Similar to a comment above, I could be OK with presenting this as an area for future research, but I think it needs to be framed by noting the limitations of this study for drawing any clinical recommendations around HAs versus haptic vibration.

Response: This was reworded: “One possibility to be tested is that individuals who use hearing aids may find haptic vibrations to be more helpful in reverberant environments when the hearing aids are removed because it would eliminate any conflict between the digital processing of the hearing aid and the vibrational signals that are providing information about the sounds of speech without processing.”

“Upon completion of the trial, the average EOC score...”: Similar to previous comments, it seems that the Results only present benefit scores but now the Discussion mentions raw EOC scores for the group with and without HAs. If raw scores are mentioned in the Discussion, they should be presented in the Results.

Response: The tables added to the supplementary material (Multimedia Appendices 2 and 3 in the final paper) will now clarify this.

This section ends by noting equivalent ending EOC scores for the group with and without HAs; a statistical result should be presented to make this claim (and should be presented in the Results section).

Response: Independent *t* test results are located in the Results section.

One additional note: Results from the final questionnaire do not seem to be presented. Is there a reason for this? Given that the APHAB is the only outcome measure, it would be beneficial to see results from the final questionnaire in this paper alongside the APHAB. The final questionnaire also measures something a little different than the APHAB—it is more holistic for the whole field trial experience.

Response: Three of our participants requested to continue use of the wristband after the study ended, and hence, they did not fill out the final questionnaire. Of those who did, some had criticisms (“I’m really unsure if the Clarify band was helpful or not”) and some had praise (“It was very beneficial. Thank you”); however, the comments were too few to be statistically meaningful. This information has been added to the Results section.

Conclusion

Same comments as before about noting that this study applies to self-perceived or self-reported benefit.

“We found that vibrotactile feedback provides more benefit for those without hearing aids than for those with hearing aids...”: From what I see in the Results section, the statistical results do not support this conclusion. The 10-point criterion from Cox cannot be applied if we are not sure the group means themselves are even different (as indicated by the insignificant P value). I think it is OK to say the data are trending in this direction and that the small n may render the study underpowered to detect this difference at P<.05. Future work is needed to establish whether this claim is true. For now, I would argue it needs to be softened based on the findings and limitations of the study design.

Response: This sentence was changed to “We found that vibrotactile feedback tends to provide more benefit for those without hearing aids than for those with hearing aids, although it does provide benefit for both. The small sample size may have rendered the study underpowered to detect this difference at $P<.05$ and further study is necessary to validate this finding.”

Finally, I suggest adding a limitations section, which could note limitations around:

- *Small n*
- *Reliance on self-report data without objective speech-testing data*
- *Potential for placebo effect to influence results*
- *Small n makes it difficult to discern whether/how individual and demographic characteristics could affect ability to integrate the haptic vibrations and benefit from the wristband—some characteristics one might wonder about include baseline cognitive ability, education level, differences in underlying degree/configuration of hearing loss, or duration of hearing loss*
- *Use of nonclinical audiogram for some participants (a minor limitation but should be noted)*
- *No information on how many hours per day the wristband was worn. One might hypothesize that outcomes could be related to wear time. Furthermore—beyond raw wear time—we also do not have information about the richness/complexity of auditory information processed through the wristband*

Response: The following Limitations section was added:

“There are limitations of this study. First, the small sample size prevents extrapolation of the results to larger populations; this will be addressed in future studies. We were also limited in our ability to collect speech comprehension data in a noise-controlled environment with standardized volume controls – this is because the testing was done in participant homes instead of a laboratory. As a result, this study depended on self-report data (APHAB) which always has the potential of being influenced by a placebo effect. Another limitation is that some participant audiograms were assessed via phone applications rather than an audiologist’s office; however, it should be noted that these appear to yield roughly equivalent results [10]. We also note that the specific type of hearing loss was also not controlled for beyond meeting the audiogram requirements. One final thing to note is that participants could move their hand (and hence their wristband), meaning that the microphone placement was not standardized in a single position. We do not consider this a limitation of the study, as the study is meant to test whether a vibrotactile wristband can be used to detect sound. The positive results reported here suggest that the mobility of the microphone does not present a problem.”

Round 2 Review

Reviewer F

General Comments

The authors appear to have responded to previous comments. However, having two different versions of the manuscript in the system has caused confusion. Having some sort of system to track changes would also have been very useful.

The authors have persisted with using Michels et al [4]; this is not a primary reference for results of noise exposure of burden of hearing loss.

Response: We have added two more additional references to support the claims of noise exposure causing hearing loss in the higher frequency ranges. Both of these references have been cited in over 100 publications:

- Chen KH, Su SB, Chen KT. An overview of occupational noise-induced hearing loss among workers: epidemiology, pathogenesis, and preventive measures. *Environ Health Prev Med*. Oct 31, 2020;25(1):65. [doi: 10.1186/s12199-020-00906-0] [Medline: 33129267]
- Hong O, Kerr MJ, Poling GL, Dhar S. Understanding and preventing noise-induced hearing loss. *Dis Mon*. Apr 2013;59(4):110-118. [doi: 10.1016/j.disamonth.2013.01.002] [Medline: 23507351]

I am not convinced that even an omnidirectional microphone would be optimally placed on the wrist.

Response: Thank you for your recommendation. The wrist placement was a decision made based on practicality for the user. In the past, we tried various form factors (including a vest), but those turned out to be impractical for daily use. During the algorithm design, different listening conditions were accounted for in the training data. In the end, our data make it clear that the current form factor works well; the future will tell if there is another form more optimal.

"...allow them to enjoy audio based entertainment such as movies and podcasts..." was of course not tested.

Response: We were not making a declaration in this sentence—we were simply identifying potential implications of improving one's ability to understand speech. ("With further development and refinement, this technology has the potential to improve the quality and productivity of their daily interactions, enable them to enjoy audio based entertainment such as movies and podcasts, help them understand conversations in complicated acoustic environments, and fill the residual gaps of impairment left by their hearing aids.")

The last paragraph of the Introduction reads like a conclusion, not the presentation of aims or objectives.

Response: We have revised the last paragraph of the introduction to now read "In this study, we aimed to demonstrate that a simple wearable sensory substitution device that transforms speech sounds into haptic vibrations on the wrist can help individuals with high frequency hearing loss to feel more confident in their ability to understand speech communication throughout their normal daily routine."

I am unconvinced about the rationale for removing aversiveness from the APHAB; the same can be said about the other subscales. It is not about the unpleasantness introduced by the device; otherwise, why should the APHAB be applied before an intervention such as HAs or cochlear implants (as done in this study)? It is the person's overall aversiveness to sound. Anyway, the data were not collected, so there is little to be done.

Response: The following questions address aversiveness in the APHAB:

- Unexpected sounds, like a smoke detector or alarm bell are uncomfortable.
- Traffic noises are too loud.
- The sounds of running water, such as a toilet or shower, are uncomfortably loud.
- The sounds of construction work are uncomfortably loud.
- The sounds of a fire engine siren close by are so loud that I need to cover my ears.
- The sound of screeching tires is uncomfortably loud.

These questions were removed because the wristband does not vibrate to any of these sounds, it only vibrates to speech sounds. These questions are completely out of context (and therefore unanswerable) for the scenarios in which the wristband would vibrate.

"What was the rationale for the specifications for the audiogram?"

"This was simply a general inclusion criterion to make certain we were capturing garden-variety presbycusis."

It would be useful for this to be mentioned.

Response: The following sentence has now been added to the manuscript: "These specifications were chosen in order to capture individuals with hearing loss profiles in alignment with high frequency hearing loss."

Figure 5 appears to be truncated at the right for day 42.

Response: Thank you for pointing this out; we have made the necessary change.

Figure 8: Why is there a –5 label for the vertical axis?

Response: There is an –5 on the vertical axis because the error bar for BN with HAs drops below the horizontal axis to −1.96

Anonymous [14]:

I appreciate the authors' thorough revision in response to reviewer feedback, and I found this version to be very much improved. It has been a pleasure reviewing this paper and learning more about the authors' interesting work on this novel device, which is now more clearly and thoroughly explained in this newest version of the paper.

I have only a few suggested minor revisions remaining as follows:

- *In the Results section of the Abstract, it says "those without hearing aids showed a 10.78 point greater drop in average APHAB benefit score at 6 weeks." I believe this should read 10.78 higher APHAB benefit score. It would be a drop in score from baseline to the 6-week score if discussing the global APHAB score, but if discussing the benefit score, then the score increased from baseline to 6 weeks.*

Response: Thank you for catching this; we have revised the sentence: "Those without hearing aids showed a 10.78 point larger improvement in average APHAB benefit score at 6 weeks than those with hearing aids."

- *In the Results section of the Abstract, most of the results are discussed as the group average, with only one result framed in terms of non-HA users versus HA users. It might be helpful to more clearly specify that when the average results are presented—it is across all participants. I do not have a strong preference on this, just something I noticed.*

Response: Thank you for the suggestion, we have reworded the Results portion of the abstract for further clarity:

"By the end of the 6 week study, the average APHAB benefit score across all participants reached 12.39 points from a baseline of 40.32 to a final score of 27.93 (SD=13.11, *n*=16, *P*=.002, two-tailed dependent t-test). Those without hearing aids showed a 10.78 point larger improvement in average APHAB benefit score at 6 weeks than those with hearing aids (t(14)=2.14, *P*=.10, two-tailed independent t-test). The average benefit score across all participants for ease of communication (EOC) was 15.44 (SD=13.88 , *n*=16, *P*<.001 , two-tailed dependent t-test). The average benefit score across all participants for background noise (BN) was 10.88 (SD=17.54, *n*=16, *P*=.03, two-tailed dependent t-test). The average benefit score across all participants for reverberation (RV) was 10.84 (SD=16.95, *n*=16, *P*=.02, two-tailed dependent t-test)."

- *In the last paragraph of the Introduction, it says "...can help individuals with high frequency hearing loss to feel more confident in their ability to understand speech communication." Although I understand why the authors are making this inference from the APHAB, it does not feel quite supported enough to jump from the APHAB results to a statement about participants' confidence. I would strongly suggest editing this to be in line with the language used throughout the rest of the paper (eg, increasing subjective assessment of speech ability, increasing self-rated communication ability, or decreasing self-perceived hearing difficulty in daily communications).*

Response: We have revised this sentence: "In this study, we aimed to demonstrate that a simple wearable sensory substitution device that transforms speech sounds into haptic vibrations on the wrist can help individuals with high frequency hearing loss perceive a greater ability to understand speech communication throughout their normal daily routine."

- *At the end of the APHAB section under Tasks, where it says "Higher benefit scores indicate...," I would also suggest adding the calculation for the benefit score as unaided – aided; then, it could be deleted from the next section.*

Response: This information is already contained in the paragraph: "The test was administered through an online questionnaire that captured the data onto a datasheet for analysis. The benefit score is calculated by subtracting the final aided score at the conclusion of the trial from the baseline unaided score that was measured at the beginning of the trial. Lower raw APHAB scores indicate lower levels of disability associated with hearing loss. Higher benefit scores indicate more perceived benefits from intervention."

- *In Table 1, I would suggest adding a column to indicate which participants had a professional hearing test and which used the app option.*

Response: Thank you for the suggestion. We have updated the table and the caption below it.

"Table 1. Demographic data. Hearing loss values are decibels of hearing loss at six pure tones in the left and the right ears. Hearing loss values are measured without cochlear implants or hearing aids. Note that 90 dB of hearing loss is the most the test can detect. Audiogram source indicates where the audiogram originated from. Audiologist indicates the audiogram was measured by an audiologist and mobile app indicates the participant provided two audiograms measured by the Mimi and Hearing & Ear Age Test Mobile apps."

- *For the Table 2 legend, I would suggest specifying how precision and recall are calculated in terms of true positives, false positives, etc. Additionally, it would be helpful to know how the F_1-score is calculated.*

Response: We have updated the caption under the table to include the equations for precision, recall, and F_1-score.

"Table 2. Algorithm performance. Precision is the ability of a classification model to return only the data points in a class. It is calculated by dividing the true positives by the sum of the true positives and false positives. Recall is the ability of a classification model to identify all data points in a relevant class. It is calculated by dividing the true positives by the sum of the true positives and false negatives. F1 Score is a single metric that combines recall and precision using the harmonic mean. It is calculated by dividing the true positives by the sum of the true positives plus one half of the sum of the false positives and false negatives."

- *In the Results section comparing non-HA users to HA users, the sentence about the 10.78-point difference could be made clearer if it specified that the non-HA users had a 10.78-point higher benefit score than the HA users (rather than just saying there is a difference).*

Response: Thank you for the suggestion, we have revised this sentence: "Results showed a 10.78 point greater APHAB benefit score at 6 weeks for participants who did not use hearing aids than for participants who did (t(14)=2.14, *P*=.10, two-tailed independent t-test, Figure 7)."

- *In the same section of the Results, it says "...average APHAB benefit over baseline..."—since the benefit score reflects a reduction in the APHAB score, I would suggest framing benefit not as being "over baseline" but rather "from baseline."*

Response: Thank you for the suggestion, we have revised this sentence: "The subgroup that did not wear hearing aids ended the study with an average APHAB benefit from baseline of 18.45 points (SD=11.70, *n*=7, *P*=.005, two-tailed dependent t-test). The subgroup that wore hearing aids ended the study with an average APHAB benefit from baseline of 7.67 points (SD=12.730, *n*=9, *P*=.11, two-tailed dependent t-test)."

- *In the Discussion section, where it says "Out of 16 participants, 14 ended the study with an APHAB score of 40 or below...." I think this would be more helpful if it said how many of them started the study with a score of 40 or*

below. I do not have a strong preference, however. Now that individual data are presented, it is much easier to contextualize the results.

Response: Thank you for the suggestion; we believe this information can be extracted from the individualized data table provided in Table 1 of the final paper.

- *In the Discussion section, it says "It is also possible that participants who started the study with a lower APHAB score had more room for improvement." I think this should say a higher APHAB score, as higher scores mean more perceived difficulty.*

Response: Thank you for the suggestion, we have revised this sentence: "It is also possible that participants who started the study with a higher APHAB score had more room for improvement, as higher APHAB scores indicate a higher degree of perceived disability."

- *In the Conclusion, it mentions that the study was underpowered to detect the difference between HA users and non-HA users at P<.05. This is presented for the first time in the Conclusion, which seems out of place. I would suggest first mentioning this in the Limitations section above. It could also be mentioned in the Conclusion, though, because it's an important point—but reading new information in the conclusion was a bit jarring.*

Response: Thank you for pointing this out. We have added this information to the Discussion section in the paragraph describing the possible differences for HA and non-HA users. There is also a mention of the failure to reach statistical significance in the Results section:

"Participants without hearing aids demonstrated a trend toward higher self-reported benefit from vibrotactile sensory substitution for speech understanding, though this did not reach statistical significance. Given that this group started the study trending toward a higher APHAB score (above), we presume the difference is because the hearing aid group already gains benefit from their technology and therefore has less room for improvement. It is difficult to predict what the interaction between hearing aids and vibrotactile feedback will be because of the differing signal processing techniques used in digital hearing aid technologies Digital hearing aids convert sound waves into numerical codes before amplifying them. This code contains information about a sound's frequency and amplitude, allowing the hearing aid to be specially programmed to amplify some frequencies more than others. Digital sound processing capabilities allow an audiologist to adjust the hearing aid to a user's needs and to different listening environments. Digital hearing aids can also be programmed to focus on sounds coming from a specific direction. It is possible the wristband represents sounds that differ significantly from those represented by the hearing aid. Future studies can possibly explore directly connecting the wristband to the user's hearing aids through a bluetooth signal so that the wristband's signals directly correspond with the sounds the user is hearing. For this study, the small sample size rendered the study underpowered to detect differences between those who used hearing aids and those who did not at $P<.05$. Future studies will be designed to further investigate this finding."

Very Minor Comments

- *First paragraph under APHAB under Tasks, suggest revising "they are asking" to "they ask"*

Response: Thank you for this suggestion, we have revised the sentence: "These questions were removed because they ask about the unpleasantness of sounds heard through a hearing aid, which does not apply for our device."

- *In the same section, suggest revising the two instances of "was referring" to "referred"*

Response: Thank you for the suggestion, we have revised this sentence as well: "If the participant regularly wore hearing aids, 'with the wristband' referred to wearing the wristband in addition to their hearing aids and 'without the wristband' referred to wearing their hearing aids alone.

- *For the Results section that discusses the BN score, it should read "16.99 points higher than those with hearing aids" ("with" is missing)*

Response: Thank you for pointing this out; we have made this correction: "The average BN benefit for those without hearing aids was 16.99 points higher than those with hearing aids (t(14)=2.14, P=.05, two-tailed independent t-test)."

References

1. Eikelboom R. Peer review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation". JMIRx Med 2024;5:e55554. [doi: 10.2196/55554]
2. Kohler I, Perrotta MV, Ferreira T, Eagleman DM. Cross-modal sensory boosting to improve high-frequency hearing loss: device development and validation. JMIRx Med 2024;5:e49969. [doi: 10.2196/49969]
3. Olusanya BO, Davis AC, Hoffman HJ. Hearing loss: rising prevalence and impact. Bull World Health Organ 2019 Oct 1;97(10):646-646A. [doi: 10.2471/BLT.19.224683] [Medline: 31656325]
4. Michels TC, Duffy MT, Rogers DJ. Hearing loss in adults: differential diagnosis and treatment. Am Fam Physician 2019 Jul 15;100(2):98-108. [Medline: 31305044]
5. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. Int J Audiol 2014 Feb;53 Suppl 1:S18-S27. [doi: 10.3109/14992027.2013.860488] [Medline: 24447233]
6. Auer ET, Bernstein LE, Sungkarat W, Singh M. Vibrotactile activation of the auditory cortices in deaf versus hearing adults. Neuroreport 2007 May 7;18(7):645-648. [doi: 10.1097/WNR.0b013e3280d943b9] [Medline: 17426591]

7. Levänen S, Jousmäki V, Hari R. Vibration-induced auditory-cortex activation in a congenitally deaf adult. Curr Biol 1998 Jul 16;8(15):869-872. [doi: 10.1016/s0960-9822(07)00348-x] [Medline: 9705933]
8. Perrotta MV, Asgeirsdottir T, Eagleman DM. Deciphering sounds through patterns of vibration on the skin. Neuroscience 2021 Mar 15;458:77-86. [doi: 10.1016/j.neuroscience.2021.01.008] [Medline: 33465416]
9. Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. Ear Hear 1995 Apr;16(2):176-186. [doi: 10.1097/00003446-199504000-00005] [Medline: 7789669]
10. Yesantharao LV, Donahue M, Smith A, Yan H, Agrawal Y. Virtual audiometric testing using smartphone mobile applications to detect hearing loss. Laryngoscope Investig Otolaryngol 2022 Dec;7(6):2002-2010. [doi: 10.1002/lio2.928] [Medline: 36544936]
11. Anonymous. Peer review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation". JMIRx Med 2024;5:e55727. [doi: 10.2196/55727]
12. Glick H, Sharma A. Cross-modal plasticity in developmental and age-related hearing loss: clinical implications. Hear Res 2017 Jan;343:191-201. [doi: 10.1016/j.heares.2016.08.012] [Medline: 27613397]
13. Hong O, Kerr MJ, Poling GL, Dhar S. Understanding and preventing noise-induced hearing loss. Dis Mon 2013 Apr;59(4):110-118. [doi: 10.1016/j.disamonth.2013.01.002] [Medline: 23507351]
14. Anonymous. Peer review of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation". JMIRx Med 2024;5:e55728. [doi: 10.2196/55728]
15. Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: an ASR corpus based on public domain Audio books. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): IEEE; 2015:5206-5210. [doi: 10.1109/ICASSP.2015.7178964]
16. Mimi Hearing Test apps. Mimi Health. URL: https://mimi.health/hearing-test-apps [accessed 2024-01-25]
17. Sakashita Y. Hearing Test & Ear Age Test. App Store. URL: https://apps.apple.com/us/app/hearing-test-ear-age-test/id1067630100 [accessed 2024-01-25]
18. Bentler RA, Niebuhr DP, Johnson TA, Flamme GA. Impact of digital labeling on outcome measures. Ear Hear 2003 Jun;24(3):215-224. [doi: 10.1097/01.AUD.0000069228.46916.92] [Medline: 12799543]

Abbreviations

APHAB: Abbreviated Profile of Hearing Aid Benefit
BN: background noise
EOC: ease of communication
HA: hearing aid

Edited by E Meinert; submitted 14.12.23; this is a non–peer-reviewed article;accepted 14.12.23; published 09.02.24.

Please cite as:
Kohler I, Perrotta MV, Ferreira T, Eagleman DM
Authors' Response to Peer Reviews of "Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation"
JMIRx Med 2024;5:e55510
URL: https://xmed.jmir.org/2024/1/e55510
doi:10.2196/55510

Authors' Response to Peer Reviews of "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020"

Thokozani Mzumara[1,2,3], MSc; Marios Kantaris[3,4], PhD; Joseph Afonne[1], OD

1

2

3

4

Corresponding Author:
Thokozani Mzumara, MSc

Related Articles:

Companion article: http://preprints.jmir.org/preprint/44381

Companion article: https://med.jmirx.org/2024/1/e58361

Companion article: https://med.jmirx.org/2024/1/e57935

Companion article: https://med.jmirx.org/2024/1/e44381

(JMIRx Med 2024;5:e57620) doi:10.2196/57620

KEYWORDS

access to health; health service utilization; eye care use; health-seeking behavior; sociodemographic determinant; visual impairment; social support; women empowerment; education; eye care; eye; ophthalmology; visual; ECU; eye service; utilization; Malawi; empowerment; health service use; use

This is the authors' response to peer-review reports for "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020."

Round 1 Review

Anonymous [1]

Potentially an interesting paper [2], but more details are needed in the methods to enable the reader to understand the results.

The English is not good throughout the document, and the writing could be much more succinct and precise.

Response: We have reviewed the language problems.

Anonymous [3]

General Comments

This paper is a secondary analysis of a Malawian household survey exploring associations of patients who self-reported as having used formal eye care services. It is a useful idea to use this survey data for this purpose, but the author needs to check that they are using the correct source numbers for their statistical analysis and only report the numbers actually surveyed—not the national estimated numbers derived from these.

Specific Comments

Major Comments

1. "In Malawi, 3.3% of the population is blind compared to 1.01% in America [4,5]." There is no way 3% of Malawi is blind. (Half of Malawi's population are children, so if 3% of Malawi was blind, that would be about 1 in 20 adults—not possible.) Check your references.

Response: Done.

2. The abstract needs improving to give the definition of eye care use (ECU). In the results, it says "The prevalence of ECU was 60.6%," which is not really a prevalence unless you give a clearer definition, for example "of those with eye symptoms, what proportion have access formal eye care services in the two weeks prior to the survey date."

Response: Done.

3. The sample was 28,388 adults? You cannot, then, in the results' "Characteristics of study participants" section say

there were 6 million young adults involved or that 5,660,836 (56%) of the adults were married. You also can't say that "27,336 (0.3% of 2,734,768) complained of ocular symptoms." This is the main problem with the report—you need to give the actual numbers of people surveyed who reported ocular symptoms—presumably 0.3% of 28,388—which is only 85 people. Thus your CIs/other statistical analyses around estimates with a sample of 85 people reporting eye symptoms will be quite different than if you extrapolate to the whole population of Malawi.

Response: We have removed the section "characteristics of the study participants."

Minor Comments

4. "We entered the variables...": Who is "we"? I only see one author.

Response: We have added other authors as initially indicated on the system database.

5. "Sort care" should be "sought care": This is used 5 times in the paper so should be changed at all uses.

Response: We have corrected the error.

6. There are some random capital letters in various places: "that In Malawi"—why has "In" got a capital?

Response: We have corrected the error.

Round 2 Review

Point-by-point response to decision letter.

Methods

1. Please justify the use of the confusion matrix in the manuscript.

Response: We have withdrawn use of the confusion matrix from the paper.

2. Please add the age of people included in the Integrated Health Survey (IHS).

3. Was this a national survey? If not, please specify the location.

Response: This was a national survey including all districts in the country.

4. Results: Please add the number of households and individuals included in the analysis.

Response: We have added the number of households and individuals included in the study

5. The sentence starting "But it was..." is not clear. Please rewrite.

Response: We have edited this part.

Main Text

Introduction

6. Most of the key elements are covered, but it is poorly written. The following paper might be another useful reference: Tafida A, Kyari F, Abdull MM, et al. Poverty and blindness in Nigeria: findings from the national survey of visual impairment and blindness. Ophthalmic Epidemiol. 2015;22(5):333-41 [doi: 10.3109/09286586.2015.1077259] [Medline: 26395660]

Response: We have rewritten the introduction and included the reference suggested above.

7. Please clarify the aims and objectives of the study. For example, results on reasons for not accessing services are presented, but this is not mentioned beforehand.

Methods (Additional Feedback)

8. To better understand the source of data used in the analysis, please explain the sample size calculation and the sampling strategy for IHS. Presumably, households are the last sampling unit; are all eligible individuals normally living in the selected households then interviewed?

Response: We have explained in the text.

9. The paragraph on sampling weights is not clear. Presumably, this reflects the sampling strategy for the IHS so that the findings can be extrapolated to the whole population aged ≥15 years? If so, please explain.

Response: We have included this explanation in the text.

10. Please outline what confusion matrix techniques are for, what the method entails, and the outputs of the analysis.

Response: We have removed this from the article.

11. What symptoms were used in the analysis?

Response: We have indicated symptoms related to eye problems.

12. "Data" is plural: should say "data were...."

Response: We have edited.

Results

13. Define young adults and older adults.

Response: We have edited.

14. I assume Table 1 shows findings extrapolated to the whole country. The results of the IHS should be presented first, followed by extrapolations to the whole country, specifying the weights used, so that the analysis is more transparent.

Response: We have presented the results of IHS without the weighting before extrapolations under Multimedia Appendix 1.

15. Table 1: Put the data into a proper table, as the results are difficult to see at the moment.

Response: We have modified Table 1 and included a write up for the the rest.

16. The results need to be better ordered. At the moment regional differences in prevalence are included in the paragraph on where participants sought care.

Response: We have rearranged.

17. I do not understand the findings in Table 3. Presumably, these are the findings from the confusion matrix technique?

Response: We have removed Table 3.

18. It would be interesting to know whether the nature of the symptoms reported influenced health-seeking behavior. For example, were those who reported loss of vision more likely to access services?

Response: This was beyond the scope of this study, as the data did not cover aspects of vision loss.

Discussion

19. Please put negligence in inverted commas, as this is a judgmental term that means failure to give enough care or attention to someone or something that you are responsible for.

Response: We have modified.

20. An important reason why chronic conditions were more common than other conditions, which are often short-lived, is because of the cross-sectional nature of the study and the short 2-week period over which participants were to report eye conditions.

Response: We have included this suggestion in the text.

References

21. Follow instructions to authors. For example, how many authors should be quoted (ref 1 in particular)?

Response: We have corrected it.

References

1. Anonymous. Peer review of "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020". JMIRx Med 2024;5:e58361. [doi: 10.2196/58361]
2. Mzumara T, Kantaris M, Afonne J. Eye care service use and associated health-seeking behaviors among Malawian adults: secondary analysis of the Malawi Fifth Integrated Household Survey 2019-2020. JMIRx Med 2024;5:e44381. [doi: 10.2196/44381]
3. Anonymous. Peer review of "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020". JMIRx Med 2024;5:e57935. [doi: 10.2196/57935]
4. Kalua K, Lindfield R, Mtupanyama M, Mtumodzi D, Msiska V. Findings from a rapid assessment of avoidable blindness (RAAB) in Southern Malawi. PLoS One 2011 Apr 25;6(4):e19226. [doi: 10.1371/journal.pone.0019226] [Medline: 21547074]
5. Flaxman AD, Wittenborn JS, Robalik T, et al. Prevalence of visual acuity loss or blindness in the US: a Bayesian meta-analysis. JAMA Ophthalmol 2021 Jul 1;139(7):717-723. [doi: 10.1001/jamaophthalmol.2021.0527] [Medline: 33983373]

Abbreviations

ECU: eye care use
IHS: Integrated Health Survey

Edited by E Meinert; submitted 21.02.24; this is a non–peer-reviewed article;accepted 21.02.24; published 09.04.24.

Please cite as:
Mzumara T, Kantaris M, Afonne J
Authors' Response to Peer Reviews of "Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020"
JMIRx Med 2024;5:e57620
URL: https://xmed.jmir.org/2024/1/e57620
doi:10.2196/57620

Author's Response to Peer Reviews of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study"

Theodore Lytras, MPH, MD, PhD

School of Medicine, European University Cyprus, Engomi, Cyprus

Corresponding Author:
Theodore Lytras, MPH, MD, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/43341

Companion article: https://med.jmirx.org/2024/1/e59639

Companion article: https://med.jmirx.org/2024/1/e59638

Companion article: https://med.jmirx.org/2024/1/e43341

(JMIRx Med 2024;5:e59637) doi:10.2196/59637

KEYWORDS

COVID-19; pandemic; health care disparities; intensive care unit; ICU; right to health; quality of care; intubation; mortality; health disparity; health inequality; surveillance data; in-patient; mortality; COVID-19 patient; hospitalization; disparity; inequality; surveillance; health care system; Greece; region; Delta; Omicron; vaccination; vaccine; public health; patient load; deterioration; time

This is the authors' response to peer-review reports for "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study."

Round 1 Review

Anonymous [1]

"We aimed to update this analysis [2] to include the large "Delta" and "Omicron" waves that affected Greece during 2021 - 2022."

So why did you analyze data also from 2020?

Response: Because we needed to consider the entirety of the relevant data (in the context of our aim to describe the relationship between patient load and in-hospital mortality), and not only the new data of 2021 and 2022. Looking at the totality of the evidence is good practice, as we are sure you would agree.

"Vaccination did not affect the mortality of these already severely ill patients."

Did you check how many months before they had received the vaccine? How many of the participants were not vaccinated?

Response: Examining the time lag between vaccination and disease onset was not part of our analysis; 11,944 of 14,011 patients (85.2%) were unvaccinated, as is already reported on Table 1; we have now included this information in the text and in the abstract.

"Vaccination did not show a statistically significant association with mortality, regardless of the number of doses received (Figure 2)."

This needs rephrasing if all participants were vaccinated at least once.

Response: No, they were not. In fact, most participants (85.2%) were unvaccinated. The described lack of association refers to our multivariable Poisson model (summarized in Figure 2), where you can see that the hazard ratios for 1, 2, or 3 doses versus zero doses are all statistically nonsignificant. Having said that, this finding is not surprising: the COVID-19 vaccine is very effective in preventing severe disease and thus death, but *once severe disease does occur*, surviving it depends on the health care received, not on vaccination.

"Between 1 September 2020 and 3 April 2022..."

I see that your previous study included data up to May 2021. Thus, there is a lot of overlap in the data of these two studies.

Response: Our current study is clearly described as being an update of the previous study (which we cite as reference 3). That study analyzed 6282 patients, whereas the current one analyzes 14,011 patients, more than twice as many. Also the original only covered the Alpha variant wave, whereas the current study extends that to the Delta and Omicron waves. It is therefore a meaningful and important update, not at all salami-sliced as you may imply.

There are many errors in the use of English and typos which makes it difficult to understand. See examples below:

"Mortality was significantly higher above 400 patients"

This is not very clear (please rephrase).

"...with an adjusted Hazard Ratio of 1.22, 95% CI: 1.09 - 1.38)"

Where does the parenthesis open?

"...rising progressively up to 1.48 (95% CI: 1.31 - 1.69) for 800+ patients."

This is not clear.

"...we found no statistically association between mortality and..."

Please correct typo.

Response: Thank you for your comment; we have rephrased/corrected all of the above.

It would be good to include DOIs to all references.

Response: Thank you, we now follow the JMIR citation style (as included in Zotero), which includes PMIDs to all relevant references.

Reviewer CF [3]

Abstract has to clarify the goal, sample, results, and health and social implications to cope with the next pandemics to improve health care.

Response: Thank you for your comment. We have extensively amended the abstract. At the same time, we are reluctant to overinterpret our findings (with respect to "social implications") especially in the context of the abstract. Still, we already make a clear reference to that in our conclusion section of the abstract ("This highlights the need for urgent strengthening of health care services in Greece, ensuring equitable and high-quality care for all").

Introduction is poor and has to better clarify the research questions of this study and provide more theoretical background about strategies of prevention and good governance to cope with the pandemic crisis. After that they can focus on the topics of this study to provide a correct analysis for fruitful discussion (see suggested readings that must be all read and used in the text).

Response: Thank you, but the aim of our study is not about general "strategies of prevention and good governance." It is specifically (and narrowly) about illustrating the association between patient load during the pandemic and in-hospital mortality of severely ill COVID-19 patients. To reflect that, our introduction is short and to-the-point, stating the context and aims of the study and setting up our methods, results and discussion.

Methods of this study are not clear. The section of Materials and Methods must be restructured with the following 3 points in the same order...

Response: Thank you for your comment. We have already structured the methods with these 3 items in the same order, just not using subheadings. The first paragraph includes the sample, data source, and variables, and the second paragraph describes the statistical analysis.

Results: It is not clear why authors apply a Mann-Whitney test, considering the large sample it is better to apply an independent sample t test, or some other parametric test. This additional test can support better results if they are reliable.

Response: With all due respect, this is incorrect. The nonparametric Mann-Whitney is used in Table 1 to compare the age distribution between patients who died and those who survived, because age does not conform to a normal distribution. It is not a matter of sample size. In any event, the point is moot, since the difference is so large that both the Mann-Whitney and 2-sample *t* test produce $P<.001$.

Table 1: avoid acronyms; specify ICU as intensive care unit. Same comments for Figure 2.

Response: We have defined the acronym ICU at first use (both in the Abstract, and in the Methods section of the manuscript) as is standard practice.

Discussion: First, authors have to synthesize the main results in a simple table to be clear for readers and then show what this study adds compared to other studies. In addition, authors should discuss the type of mechanical ventilation, because studies show that invasive ventilation (intubation) creates VAP (ventilator-associated pneumonia), and a lot of people intubated for COVID-19 died from this problem rather than COVID-19. Countries that have reduced mortality, such as Germany and New Zealand, used mainly noninvasive ventilation, which can better treat patients and avoid mortality with new technology.

Response: Thank you, but we have already summarized all the main results in Figure 2 (which includes all effect sizes and 95% CIs). All our patients were mechanically ventilated; therefore, we do not have empirical data to discuss the point that the reviewer suggests. Thus, if we were to do so, it would be pure speculation.

Conclusion has to be added as an autonomous section. Conclusion not to be a summary, but authors have to focus on manifold limitations of this study and provide suggestions for health, crisis management, and social policy, as well as how nations can prevent, with good governance and new technology in artificial ventilation, the next pandemics and improve health care with vaccination, noninvasive ventilation, and nonpharmaceutical measures of control. In this manner the paper can provide useful policy implications for Greece and improve health care for the next pandemics.

Response: Thank you, but this is a short communication and—more importantly—we want to avoid overinterpreting our data or venturing beyond what our data can support (which is what the reviewer suggests). May we also point out that our original analysis (of which this current study is an update) became hotly debated in Greece, which is another reason we wish to be cautious in our interpretation. Nevertheless, the very last sentence of our Discussion section clearly and unambiguously offers our conclusion and policy implications: "The findings highlight the need for urgent strengthening of health care services in Greece in order to improve their performance and ensure equitable access to high-quality care for all."

References

1. Anonymous. Peer review of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study". JMIRx Med 2024;5:e59639. [doi: 10.2196/59639]
2. Lytras T. Health care system overstretch and in-hospital mortality of intubated patients with COVID-19 in Greece from September 2020 to April 2022: updated retrospective cohort study. JMIRx Med 2024;5:e43341. [doi: 10.2196/43341]
3. Coccia M. Peer review of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study". JMIRx Med 2024;5:e59638. [doi: 10.2196/59638]

Edited by E Meinert; submitted 17.04.24; this is a non–peer-reviewed article;accepted 17.04.24; published 10.06.24.

Please cite as:
Lytras T
Author's Response to Peer Reviews of "Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study"
JMIRx Med 2024;5:e59637
URL: https://xmed.jmir.org/2024/1/e59637
doi:10.2196/59637

Authors' Response to Peer Reviews of "Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study"

Mara Giavina-Bianchi, MD, PhD; Edson Amaro Jr, MD, PhD; Birajara Soares Machado, PhD

Big Data Department, Hospital Israelita Albert Einstein, Sao Paulo, Brazil

Corresponding Author:
Mara Giavina-Bianchi, MD, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/50803

Companion article: https://med.jmirx.org/2024/1/e56496

Companion article: https://med.jmirx.org/2024/1/e56446

Companion article: https://med.jmirx.org/2024/1/e50803

(JMIRx Med 2024;5:e56441) doi:10.2196/56441

KEYWORDS

artificial intelligence; adoption; acceptance; opinion; perceptions; survey; expectations; physician; medical survey; qualitative study

This is the authors' response to peer-review reports for "Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study."

Round 1 Review

Dear Reviewers,

We deeply appreciate your time and effort to analyze our manuscript and make suggestions to improve it. We believe it was very helpful and practically all of them were incorporated in this new version. Thank you very much!

Reviewer AE [1]

General Comments

The manuscript [2] delves into the perspectives of Brazilian physicians on the integration of artificial intelligence (AI) in medical practices through an online cross-sectional survey.

Specific Comments

Major Comments

1. The study purports to evaluate the acceptance of AI by physicians, but the specific types of AI technologies explored remain ambiguous. Are they examining generative AI, natural language processing tools, classical machine learning or other uses of AI?

Response: There was not a specific technology evaluated, but, most of the times, we were exploring the idea of having a computer-aided solution, such as an AI algorithm, to aid physicians in clinical practice to diagnose, manage, or interpret exams. We introduced this sentence to clarify it in "Methods": "No specific technology was evaluated. Most of the questions asked about the use of AI algorithms for diagnoses or management of diseases, aiming to address the possible expectations of our target population composed by physicians in clinical practice."

2. The phrase "Although scarcely used in real practice" comes across as too assertive. Consider a softer phrasing.

Response: We modified it to "Although not so frequently used in daily practice yet."

3. The methods section should provide a more comprehensive breakdown of the questionnaire's design process. Which question types were chosen (Likert scale, yes or no, or numerical), and for what reasons?

Response: Thank you for this suggestion! We modified the "Methods" section to better explain the questionnaire, such as:

"The questionnaire was divided into five sections. The first one was the Informed Consent Term (question 1). Section 2 (questions 2-12) was designed to profile the physicians (sex, age, highest level of education, medical specialty, years since graduation, private versus public sector work, city and state of work, self-assessment knowledge of AI in general, and use of AI solutions in general for daily tasks. Section 3 (questions 13-18) was thought to explore the physicians' thoughts about AI solutions for diagnosis, management, subsidiary exams interpretation of diseases, as COVID-19, for example, and about

the use of AI solutions for diagnosis or treatment of diseases by nurses, physiotherapists or directly by the patient. We also proposed a hypothetical exercise to evaluate physicians' anxiety feelings and actions taken if him/herself had received a suspicious diagnose of melanoma for one of his or her skin lesions by an AI algorithm. Section 4 (questions 19-24) asked about expected benefits and problems, possible frequency of AI adoption, workload, and utility. Section 5 (questions 25-30) are about physicians' replacement by AI solutions, financial expectations, possible scenarios of AI and physicians' disagreements, legal and regulatory aspects. Along with the questions, there were many opportunities for physicians to make comments in an open box about the answers. Physicians could skip to answer any question, thus number of responders could vary along the questionnaire."

Also, the supplementary file, with the complete questionnaire, was uploaded.

4. When presenting results, always give raw data (numerator/denominator) along with percentages, especially after statements such as "Most of them described their AI knowledge as intermediate."

Response: Thank you, we modified all results and the abstract to give raw data and their percentages.

5. There is a noticeable omission of a power analysis. How can we ascertain that the sample size sufficiently represents the broader population? The description of the target population needs elaboration.

Response: As we performed a survey, obviously, there were many questions to be answered by the physicians, but we picked question 21 as the most significant to compare between groups with ≤20 years and >20 years since graduation and to estimate our power level analysis. Question 21 asked how frequent the physicians would adopt AI solutions in their daily practice if they were proven to be reliable and took up to 2 minutes of their time. The possible answers were never, rarely, and sometimes (which we grouped as not favorable) and frequently, most of the times, and always (grouped as a favorable opinion). We estimated that we would have around 35% of favorable responses in the group with >20 years since graduation and 65% in the group with ≤20 years since graduation. Considering an α error of .05 and a β error of .20, we calculated that 48 participants in each group would give us an 80% power level. As we had 163 participants, 103 with >20 years and 60 with ≤20 years, we reached the proposed power level.

However, one other reviewer suggested to divide the years since graduation in 3 groups: ≤10 years, 11-20 years, and >20 years, and repeat the statistical analysis. We did it, and we found that there is a significant difference among the 3 groups (*P*=.0402), as we can see in Table 3.

Our target population was made of physicians who were part of the clinical staff of the Hospital Israelita Albert Einstein and is not intended to represent the entire population of Brazilian physicians. The intention was to point out that the research was done with Brazilian physicians and not to state that the physicians from our hospital represent the Brazilian physician population, because they do not. The aim was to highlight the fact that this was a study conducted in Brazil, which is a geographic region of the world not yet studied in the subject. To clarify that, we changed some of the words, such as from "Brazilian physicians" to "physicians from Brazil," in the "Introduction." We included this sentence in "Methods": "Although physicians of HIAE do not represent the entire population of physicians in Brazil, their answers can give some insights of the subject, since, at this moment, we have none," and added this sentence in the "Discussion": "Our target population does not intend to represent the entire population of Brazilian physicians. Even so, a survey performed in one single, large, private hospital can be a way of drawing attention and start a debate about this new subject in Medicine among our physicians, besides capturing their expectations on the topic."

Brazil is a very large and heterogeneous country, including its health system, and 85% of its 210 million inhabitants rely on public health, which has many more issues than private ones.

Minor Comments

1. The statement "Artificial intelligence (AI) applied to Medicine has been a trending subject in recent years" is preferable over mentioning it as the "hottest topic."

Response: We modified it to "Artificial intelligence (AI) applied to Medicine has been a trending subject for the past years."

2. In Table 1, the age bracket should read "50-65" as the "50" seems to be missing. Several P values appear without context, for instance: "10. General AI Knowledge (n=164); P=.2565," "11. Regularity of AI tool usage in daily life (n=164); P=.9792," and "12. Familiar with medical AI solutions? (n=164); P=.2774."

Response: Thank you for pointing out our mistake. All tables were modified as a suggestion of the other reviewer, and a new table was incorporated into the manuscript. The significant *P* values were explained in the footnotes. *P* values that are not significant were removed from the tables.

In "Methods," we try to clarify our tests, such as the following: "There were two different questions involving statistical analysis. In the first one: 'does the time since medical graduation matter?,' the subjects were divided into 3 groups ≤10 years, 11-20 years, or >20 years, according to the answer to question 7 of the questionnaire. In the second, 'does the sex matter?,' physicians were divided in male or female, following their answers in question number 3. Statistical analyses between for both analyses were performed using the χ^2 test in Prism software version 6 (GraphPad Software, Inc., San Diego, CA, USA). *P* value <.05 was considered significant."

Also, the supplementary file, with the complete questionnaire, was uploaded, which can give a better context of the questions.

Tables were also modified to try to improve the questions' understanding.

3. Tables 2 and 3 also contain P values that require explanations or clarifications.

Response: See above.

Reviewer AJ [3]

General Comments

This paper reports the results of a survey on medical expectations on artificial intelligence solutions in daily practice. The authors argue that it is important to know the opinion that physicians would have as users of these solutions, and the reviewer could not agree more. Therefore, the results of this work may be of interest to the community.

Specific Comments

Major Comments

1. The authors say that these results represent the opinion of Brazilian physicians. Perhaps that is a bit presumptuous, at least without somehow justifying the size of the hospital relative to the Brazilian population. What percentage of the Brazilian population attends this hospital? What percentage of Brazilian physicians works there?

Response: The intention was to point out that the research was done with Brazilian physicians and not to state that the physicians from our hospital represent the Brazilian physician population, because they do not. The aim was to highlight the fact that this was a study conducted in Brazil, which is a geographic region of the world not yet studied in the subject. To clarify that, we changed some of the words, such as from "Brazilian physicians" to "physicians from Brazil," in the "Introduction," and we included this sentence in "Methods": "Although physicians of HIAE do not represent the entire population of physicians in Brazil, their answers can give some insights of the subject, since, at this moment, we have none."

Brazil is a very large and heterogeneous country, including its health system, and 85% of its 210 million inhabitants rely on public health, which has many more issues than private ones.

2. I have not been able to find the supplementary material anywhere. Therefore, I could not review the complete questionnaire.

Response: We provided the complete questionnaire as a supplementary file. I believe now you will be able to access it. Sorry for that!

3. The division into <20 years of practice and >20 years of practice does not seem sufficient to this reviewer, since in <20 years of practice you can still have quite senior physicians. I would add an additional division:<10 years, 10-20 years, and >20 years of practice.

Response: We appreciated your thoughts! We followed your suggestion and incorporated the results into the tables. We found that there was a significant difference in the frequency of intention to use AI solutions according to this new division of years since graduation. Physicians with ≤10 years since graduation are more prone to use it always or most of the times than those with >20 years since graduation, as you will see.

Minor Comments

4. How are the percentages calculated in Table 1? The percentages of every column should sum up to 100.

Response: There was a missing line in Table 1. Anyhow, the percentages were always calculated within the given column.

5. Could the authors comment on, if the physicians reported it in the questionnaire, which AI solutions they used in their daily life? Are they used in their personal life or in their work?

Response: No, the specific app (which uses AI algorithms) in their daily lives was not asked, but we believe it is the same as most of the people in Brazil: Instagram, WhatsApp, Waze, Google Apps, Alexa, Siri, Twitter, and banks app.

We did not ask about algorithms already used in medical practice, because we knew that they were only available for a small part of the target population, such as radiologists, in our hospital.

6. I assume there is an issue with the color legend for "Work facilitation" in Figure 2.

Response: Indeed! Thank you again for pointing that out! It was corrected.

7. I would not only say that physicians think AI will not interfere with the number of appointments. A third of them thinks that AI solutions will increase the number of appointments.

Response: Agreed! We highlighted that in the results part.

8. I would include, if possible, a subanalysis of the responses per gender and discuss if there is any differences.

Response: Very nice suggestion! It was done. We found one statistical difference regarding the acceptance of AI use by other hospital staff (nurses and physiotherapists). Female physicians were more favorable to it than male physicians (P=.0079).

Thank you both again!

Best regards.

Round 2 Review

Dear Reviewer AJ and Editor,

Thank you again for your answer and suggestions to improve our manuscript.

We hope we can address those comments properly.

If there is any other comments or suggestions, please, let us know.

Many thanks again and best regards.

Reviewer AJ

General Comments

This reviewer thanks the authors for the work done to improve the quality of the paper with this revision. However, I still have some comments.

Specific Comments

Major Comments

1. In the previous review round, I asked about the AI solutions the health care workers used in their daily life.

Response: We did not ask about the use of specific AI solutions for health care workers in our questionnaire, because we only have a few options, and they are very limited in range and not widespread among specialties. Therefore, we thought only few physicians had already had any contact with AI solutions in our hospital. It is not our reality yet. So, all questions were asked as a hypothetical. Due to this reality, we asked question 12:

"12) Are you aware of any ARTIFICIAL INTELLIGENCE algorithms that have been approved for medical use?

Yes

No

I am not sure"

Anyway, in my opinion, if we would redo the research nowadays, I think that question would be very interesting to pose.

The authors replied by saying "the specific app (which uses AI algorithms) in their daily lives was not asked, but we believe it is the same as most of the people in Brazil: Instagram, WhatsApp, Waze, Google Apps, Alexa, Siri, Twitter and banks app."

Response: This comment was made because we understood that you wanted to know which apps were asked for in question 11, which was not related to medicine or health care:

"11) If there is an option to use ARTIFICIAL INTELLIGENCE for some task in your day-to-day life, how often would you choose to use them?

Never

Rarely

Sometimes

Often

Always

I do not know"

This reviewer thinks this should be commented somewhere in the manuscript. From this questionnaire question, it seemed that workers have access to true AI solutions in their daily lives. However, these apps the authors mentioned as "AI solutions" use AI in their workflow but are not entirely based on AI and should not be considered "AI solutions." Without commenting on this, the reader may think that the experience of this population in the use of AI is greater than it really is.

Response: Thank you. We agree with you. We modified the paragraph in "Methods" that explains Section 2 of the questionnaire to the following:

"Section 2 (questions 2-12) was designed to profile the physicians (sex, age, highest level of education, medical specialty, years since graduation, private versus public sector work, city and state of work, self-assessment knowledge of AI in general (not specific for health care AI solutions), and use of computer or smartphone applications that use AI solutions for daily tasks, such as WhatsApp, Instagram, Facebook, Waze, Google Map, Bank's app, among others. We did not ask specifically about the use of IA solutions for healthcare in their daily work, only if they were aware of AI solutions in Medicine."

We did not ask specifically about the use of AI solutions for health care in their daily work, for the reasons explained above.

We also modified a sentence in "Descriptive Results" to mark the difference between AI solutions in general and AI solutions for health care workers:

"Most of the participants use smartphones or computers applications that incorporate AI algorithms for daily tasks outside of work (119/164, 73%) and claim to be aware of AI algorithms applied specific to Medicine (86/164, 52%)."

We also included a sentence in the "Discussion": "Even so, a survey performed in one single, large, private hospital can be a way of drawing attention and start a debate about this new subject in Medicine among our physicians, besides capturing their expectations on the topic, as AI solutions for healthcare workers are only few, very limited in range and not widespread among specialties in our reality."

Minor Comments

2. I have not yet been able to access the supplementary material, and the color legend in Figure 2 is still not fixed.

Response: I will contact JMIRx support to question why you cannot access our supplementary material, once our new Figure 2 is uploaded in the system.

Figure 2 was fixed last time. We uploaded a new version with only 2 colors as we only had 2 answers for "work facilitation": "yes" and "not alters." There was no "no" answers, so we did not include "no" in the figure. I hope you can also access both, since we find them necessary for your assessment.

3. In the text, it appears as P=.079, which is not significant. Please check.

Response: Thanks again. We corrected the *P* value to .0079 in the text, as it appears in the Table 2 footnote.

4. The P=.0513 in Table 2 is not significant.

Response: Yes, we deleted this *P* value in Table 2.

5. There should be a "Total" column in Table 1.

Response: We inserted a "TOTAL column n/(%)" as asked. Thank you.

References

1. Baglivo F. Peer review of "Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study". JMIRx Med 2024;5:e56496. [doi: 10.2196/56496]
2. Giavina-Bianchi M, Amaro Jr E, Machado BS. Medical expectations of physicians on AI solutions in daily practice: cross-sectional survey study. JMIRx Med 2024;5:e50803. [doi: 10.2196/50803]
3. Carrasco-Ribelles L. Peer review of "Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study". JMIRx Med 2024;5:e56446. [doi: 10.2196/56446]

Abbreviations

AI: artificial intelligence

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 25.03.24.

Please cite as:
Giavina-Bianchi M, Amaro Jr E, Machado BS
Authors' Response to Peer Reviews of "Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study"
JMIRx Med 2024;5:e56441
URL: https://xmed.jmir.org/2024/1/e56441
doi:10.2196/56441

Authors' Response to Peer Reviews of "Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis"

Tigran Makunts[1], PharmD; Haroutyun Joulfayan[2]; Ruben Abagyan[1], PhD

1

2

Corresponding Author:
Ruben Abagyan, PhD

Related Articles:

Companion article: https://preprints.jmir.org/preprint/55976

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.19.23298750v1

Companion article: https://med.jmirx.org/2024/1/e59120

Companion article: https://med.jmirx.org/2024/1/e55976

(JMIRx Med 2024;5:e58273) doi:10.2196/58273

KEYWORDS

monotherapy; GLP; FDA; adverse; reporting; thyroid; cancer; oncology; cancers; neoplasm; neoplasms; drug; drugs; pharmacy; pharmacies; pharmacology; pharmacotherapy; pharmaceutic; pharmaceutics; pharmaceuticals; pharmaceutical; medication; medications; glucagon-like peptide-1; Food and Drug Administration

This is the authors' response to peer-review reports for "Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis."

Round 1 Review

Anonymous [1]

General Comments

In this manuscript titled "Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis" [2], the authors analyzed over 18 million reports from the Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS), among which over 17,000 cases were identified to have increased possibility of thyroid hyperplasia and neoplasm when taking glucagon-like peptide-1 (GLP-1) receptor agonist (RA) monotherapy. The data were compared to the cases where the patients were taking sodium-glucose cotransporter-2 (SGLT-2) inhibitor monotherapy. Please see my suggestions and concerns below.

Response: We thank the reviewer for their thoughtful and constructive comments. Please see each of the concerns addressed below:

Suggestions and Concerns

1. The authors compared data between GLP-1 RA monotherapy and SGLT-2 inhibitor monotherapy. It would be great to have a little bit more introduction or description for the readers to understand why they compared the data with SGLT-2 inhibitor monotherapy? What is the function of SGLT-2 inhibitors during the therapy? I realized that the authors described this a bit in the Methods section, but more information would be appreciated to be added in the Introduction section for the readers to understand the rationale.

Response: Thank you for the suggestion. The *Introduction* section has been expanded with more details on SGLT-2 inhibitor's mechanism of action and its place in diabetes management guidelines.

2. In Table 1, how are the unique individual thyroid hyperplasia and/or thyroid neoplasm–related adverse event (AE) reports being counted? Was it the sum of the cases number searched by the above AE Preferred Term in each section, or it was counted through searching a specific AE Preferred Term?

Response: Since there were cases in FAERS where multiple AEs occurred in the same report, we made sure that there was no overcount of the AEs by counting the unique cases rather than the total number of AE Preferred Terms. Table 1's caption was expanded to clarify the definition.

3. In the reporting odds ratio (ROR) analysis, where ROR = (a/b) / (c/d), I assume a indicates the number of AE cases in the exposed group, b indicates the number of non-AE cases in the exposed group, c indicates the number of AE cases in the control group, and d indicates the number of non-AE cases in control group. Is this correct? For calculating the ROR for all GLP-1 RAs (n=17,653; number of AEs=191) compared to the SGLT-2 inhibitor (n=14,102; number of AEs=7), maybe I am wrong, but should the ROR = (191 / [17,653−191]) / (7 / [14,102 − 7]), where (17,653 − 191) and (14,102 − 7) are the non-AE cases, which equals to 22.02? Did the authors use all the cases instead of the non-AE cases for the calculations? The same applies to the other ROR numbers and 95% CIs.

Response: Yes, that is correct. Relative risk was calculated in error, instead of odds ratios. Thank you for catching the error! The numbers have been corrected in text and in the figure.

4. The authors claimed that GLP-1 RA monotherapy reports manifested a statistically significant increase in thyroid hyperplasia and neoplasm AEs when compared to SGLT-2 inhibitors. How was the statistical significance determined? Was it because the calculated ROR is over 1 (or 10) or the interval (ROR [−], ROR [+]) is large?

Response: Significance was determined by the ROR and the lower bound of the 95% CI being above 1.

5. Figure 1's resolution seems low in the document.

Response: Thank you for the suggestion. A higher-resolution file has been uploaded to the journal.

References

1. Anonymous. Peer review of "Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis". JMIRx Med 2024;5:e59120. [doi: 10.2196/59120]
2. Makunts T, Joulfayan H, Abagyan R. Thyroid hyperplasia and neoplasm adverse events associated with glucagon-like peptide-1 receptor agonists in the Food and Drug Administration Adverse Event Reporting System: retrospective analysis. JMIRx Med 2024;5:e55976. [doi: 10.2196/55976]

Abbreviations

AE: adverse event
FAERS: Food and Drug Administration Adverse Event Reporting System
FDA: Food and Drug Administration
GLP-1: glucagon-like peptide-1
RA: receptor agonist
ROR: reporting odds ratio
SGLT-2: sodium-glucose cotransporter-2

Edited by E Meinert; submitted 11.03.24; this is a non–peer-reviewed article;accepted 11.03.24; published 01.05.24.

Please cite as:
Makunts T, Joulfayan H, Abagyan R
Authors' Response to Peer Reviews of "Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis"
JMIRx Med 2024;5:e58273
URL: https://xmed.jmir.org/2024/1/e58273
doi:10.2196/58273

Authors' Response to Peer Reviews of "Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018"

Ali Hazim Mustafa[1,*], MBChB, HD CM; Hanan Abdulghafoor Khaleel[2,*], MBChB, MSc CM, DrPH; Faris Lami[3,*], MBChB, PhD

[1]

[2]

[3]

[*]all authors contributed equally

Corresponding Author:
Hanan Abdulghafoor Khaleel, MBChB, MSc CM, DrPH

Related Articles:

Companion article: https://preprints.jmir.org/preprint/54611

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.13.23298496v1

Companion article: https://med.jmirx.org/2024/1/e60433

Companion article: https://med.jmirx.org/2024/1/e60394

Companion article: https://med.jmirx.org/2024/1/e54611

(JMIRx Med 2024;5:e60194) doi:10.2196/60194

KEYWORDS

human brucellosis; livestock; clustering; spatial; temporal; Iraq

This is the authors' response to peer-review reports for "Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018."

Round 1 Review

Anonymous [1]

General Comments

This paper [2] presents a spatiotemporal distribution analysis of the outbreak of the brucellosis in Iraq from 2007 to 2018, providing explanations for potential underlying causes. The methods employed include descriptive analysis and Getis-Ord G_i. The paper exhibits a well-structured format, clear language, rich content, and appropriate methodology.*

Response: Thanks for the description.

Specific Comments

Major Comments

*1. The Abstract and the main text exhibit inconsistency in describing the methods employed. The Results section of the main text only includes the results of the descriptive analysis and Getis-Ord G_i *, with no mention of the Moran I method as indicated in the Abstract.*

2. The methods used in the paper should be briefly explained in the Methods section to clarify their principles.

3. In the Results section, the authors state that there is an increasing trend in female cases from 2016 onward. This conclusion cannot be drawn; female cases increased from 2016 to 2017 and then decreased by 2018, falling below the 2016 quantity.

4. Include spatial distribution maps of the incidence rates for 1-2 years during the study period.

Response: Maps of 2013 and 2018 were added.

Minor Comments

5. Add numerical labels to the bars in Figure 1 for a more intuitive understanding.

Response: The numerical labels were added.

6. Figure 4 lacks coordinate axes, and there is an incomplete gray box on the horizontal axis, affecting aesthetics.

Response: Figure 4 was revised.

7. Please provide the formula for calculating the case frequency.

Response: Thanks for the note. I have added the formula for calculating the incidence per 100,000 as there was no relevant

formula for the case frequency—the ones available are math related.

Anonymous [3]

General Comments

This paper talks about human brucellosis in Iraq and bring an interesting spatiotemporal analysis of the human cases in the country. The paper will contribute to the understanding of human brucellosis in Iraq and can be one more example of the use of spatiotemporal analysis for the control of the disease. However, some changes need to be made to clarify some information in the paper.

Response: Thanks for the description.

Specific Comments

Major Comments

Introduction

1. References are missing in the second phrase of the second paragraph.

Response: Thanks for noticing that. The reference was added.

2. In the third phrase, "(dogs)" is not necessary.

Response: Corrected.

3. The breeding season of sheep and goats is not in spring. So, the phrase "However, it usually coincides with the livestock breeding season, spring" should be changed, as well as "Human exposure to livestock or their contaminated products will occur during spring." The second part of this phrase is true but not absolute, since animal products can travel to other places; so for clarity, I think this phrase should be reformulated.

Response: We rewrote the sentence according to the note. However, in Iraq and other parts of Asia, breeding and lambing occur mostly during the spring months (from March to May).

4. Please, develop this paragraph further: "A study from northern Iraq showed that the prevalence of brucellosis in livestock varied from 1% to 70%, depending on the species and diagnostic methods. 4 Veterinary vaccination program started in 2007. However, its implementation was negatively affected by insecurities in the region.4." It would be very interesting to know more about the testing and the insecurities of the population regarding vaccination.

Response: Thanks for the feedback. The prevalence statement was explained further. In addition, the effect of insecurities was also explained.

5. Please connect more the idea of the first phrase of the fourth paragraph with the following ideas.

6. Please define "MOH's" in the last paragraph.

Response: Defined.

Methods

7. Please make it clear throughout that the data are about humans and not animals.

Response: Thanks for the note. We made sure to refer to human brucellosis in all the paragraphs and throughout the manuscript.

8. In the data description, why does the data come from different sources and with different types of organization and grouping? Can you make it clearer?

Response: Human brucellosis cases were retrieved from the Surveillance Section at the Ministry of Health. The data were aggregated at the provincial level until 2012. Thereafter, the data became aggregated at the district level.

9. What is HB? Please define it before using the abbreviation.

Response: It refers to human brucellosis. We spelled it completely instead of using HB.

10. What is period one?

Response: The entire phrase was changed to reflect how data collection at the Surveillance Section changed and how it affected the analysis of the data. The study period from 2007 to 2018 was divided into 2 parts: the first one spanned from 2007 to 2012 when the data were aggregated at the provincial level, and the second part spanned from 2013 to 2018 when the study data were aggregated at the district level.

11. What does "the low or high attribute zone" mean?

Response: It means clusters of districts with low or high incidence. The statement was changed to make it clearer.

12. Please explain the P values assumed for the "Getis-Ord Gi statistic" and what exactly this statistic is identifying.*

Response: Thanks for the feedback. A description of the *P* values, its meaning, and interpretation was added to the *Methods*.

13. In the Abstract, the statistical analysis is explained differently from the Methods. Please make them similar.

Response: Thanks for the note. The necessary edits were made.

Discussion

14. Have you tested the trend? If so, please clarify the methods and results, but if not, please reformulate the phrase where the word is being used to describe the occurrence of your data.

Response: Thanks. We reformulated the sentence.

15. Can you define what were the "ISIL events"?

Response: It refers to the Islamic State in Iraq and the Levant.

16. "The number of females has been constantly higher than males." You are talking about humans, please used woman and man and clarify, in the Methods section, whether this classification was self-made or not.

Response: Thanks for the constructive feedback. Edits were made in the *Discussion*, *Methods*, and figure.

17. "housekeeping and farming activities. 5 (11, 12,13)." I did not understand the different configuration of the references in here.

Response: Corrected.

18. "However, this age category was very broad and could have been classified into two to three age groups to detect the most commonly affected age group. (14-17)." Why was this change not made? Again, the references are in a strange configuration.

Response: This change was not made because the data were collected in an aggregated format in these categories and not as a continuous variable that we can regroup.

19. "The infected animals must be eradicated by slaughtering and burning because there is no curable medical therapy for animal brucellosis." I did not understand this phrase. Please remove the word eradication from here and explain better why you are saying that the animals should be burned. I do understand that positive animals should be slaughter and their carcasses should be disposed of in the right way, but I have not heard of burning before.

Response: Thanks for the constructive feedback. Changes were made to reflect the appropriate control methods in animals.

20. "On the other hand, humans may consume this infected milk unpasteurized, resulting in infection and areas endemic with brucellosis to animals and humans." Change "to" for "in."

Response: Corrected.

21. There is an "18" in paragraph 5 that could be a reference. I did not understand the phrases that followed the 18.

Response: We corrected the reference. We changed it to clarify the meaning.

22. The paragraph "Transboundary transfer of animal brucellosis in the region from the neighboring countries such as Iran, Turkey, and Syria were provoked by war and political instability, lack of immunization and animal quarantine, frequent trading, low awareness and poor knowledge of HB prevention and control, residents with poor sanitary conditions easily exposed to Brucella contaminated food and water sources." is disconnected from the rest of the text; I did not understand what exactly this is about.

Response: The entire paragraph was deleted.

23. The first and second periods of the study are not clear for me.

Response: This sentence was clarified in the *Methods*.

Conclusion

24. The last part of the conclusion would be better in the Discussion section, such as "Preventive measures such as health education activities should be performed in high-risk areas. Adopting the Quarantine-Slaughter-Immunization strategy and One Health Approach is crucial in controlling the disease. This can be achieved through multisectoral coordination and coordination with neighboring countries in the control programs."

Response: Changed.

Round 2 Review

Anonymous [1]

Specific Comments

Major Comments

1. Maybe I did not express it clearly, but for the local Getis-Ord G_i method, which is one of the main methods applied in this paper, the authors should give the formula for its calculation and add the source.*

Response: Both the formula and the reference were added.

2. This is not a comment that has to be revised. Generally, the significance and spatial location of clusters in the local Getis-Ord G_i results are shown on the same map; for example, hot spots with different levels of significance are represented by 3 progressively deeper red colors, and cold spots with different levels of significance are represented by 3 progressively deeper blue colors. Also, Figure 5 contains too many maps, and it is more concise to show the results for 1 year in 1 map.*

Response: Thanks for the clarification. We cannot just show the results for 1 map as we are interested in displaying changes over time.

3. The elements that are really necessary inside a map, including but not limited to a scale, a compass, and preferably the addition of national boundaries, are missing.

Response: Totally true. However, for the purpose of this study, we are interested in the spatial distribution of human brucellosis and how it changed over time. Adding or removing other complementary features will not affect the results, and adding them may negatively affect the clarity of the map. The borders are, however, displayed.

Other Comments

The authors have finished revising, and I do not have any questions.

Response: Thanks for your useful feedback.

Anonymous [3]

General Comments

This paper brings important information and analysis on human brucellosis in Iraq. To improve the paper's understanding, I suggest an English review of the paper to improve the writing of the paper.

Response: Thank you for the useful feedback.

Specific Comments

Major comments

1. Abstract, section Methods: "The trend of cases by sex and age group were displayed from 2007 - 2018 were displayed." Please delete the last "were displayed."

Response: Corrected.

2. "The seasonal distribution of the cases from 2007 to 2012 was graphed." Substitute "was" for "were."

Response: Corrected.

3. Introduction: The paragraph on the percentages of occurrence of brucellosis only present the values but does not make a value judgment or interpret what these values mean, why are they important, and so on. Please, reformulate again.

Response: An explanation was added.

4. Discussion: Second paragraph: Make it clear that the number of woman you are talking about is the number of woman positive for brucellosis among the analyzed years.

Response: Corrected.

5. Conclusion: Substitute HB for human brucellosis.

Response: Corrected.

References

1. Anonymous. Peer review of "Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018". JMIRx Med 2024;5:e60433. [doi: 10.2196/60433]
2. Mustafa AH, Khaleel HA, Lami F. Human brucellosis in Iraq: spatiotemporal data analysis from 2007-2018. JMIRx Med 2024;5:e54611. [doi: 10.2196/54611]
3. Anonymous. Peer review of "Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018". JMIRx Med 2024;5:e60394. [doi: 10.2196/60394]

Edited by E Meinert; submitted 03.05.24; this is a non–peer-reviewed article;accepted 03.05.24; published 03.07.24.

Please cite as:
Mustafa AH, Khaleel HA, Lami F
Authors' Response to Peer Reviews of "Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018"
JMIRx Med 2024;5:e60194
URL: https://xmed.jmir.org/2024/1/e60194
doi:10.2196/60194

Author's Response to Peer Reviews of "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study"

Tatsuki Onishi[1,2,3], MPH, MD

1

2

3

Corresponding Author:
Tatsuki Onishi, MPH, MD

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.01.13.24301276v1

Companion article: https://med.jmirx.org/2024/1/e63862

Companion article: https://med.jmirx.org/2024/1/e56405

(JMIRx Med 2024;5:e63859) doi:10.2196/63859

KEYWORDS

diabetes mellitus; gastrectomy; gastric cancer surgery; glucose metabolism; postoperative diabetes onset; surgery outcomes

This is the author's response to peer-review reports for "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study."

Round 1 Review [1]

This review is the result of a virtual, collaborative live review discussion organized and hosted by PREreview and JMIR Publications. The discussion was joined by 18 people: 2 facilitators, 4 members of the JMIR Publication team, and 12 live review participants. Konstantinos Georgiou, Maria Florencia Grande Ratti, and Naser Kamyari wished to be recognized for their participation in the live review discussion, even though they have not contributed to authoring the review below. We thank all participants who contributed to the discussion and made it possible for us to provide feedback on this preprint.

Summary

This study [2] compares the results of Roux-en-Y (RY) reconstruction with other surgical techniques (OT) to determine the incidence of postoperative diabetes in patients with gastric cancer who had undergone total gastrectomy. The Tokyo Metropolitan Bokutoh Hospital cohort of 715 patients from 2005 to 2019 was examined. The study finds a statistically significant difference in the incidence of postoperative diabetes between the RY and OT groups, with RY associated with a greater incidence, through careful data preprocessing and statistical analysis. The study does admit many limitations, though, such as the absence of a control group that did not undergo a gastric bypass and the lack of assessment of the role that lifestyle factors and genetic predisposition play in the development of diabetes. The study also suggests more investigation into the possible effects of laparoscopic jejunal interposition reconstruction on gut flora and postoperative outcomes.

This retrospective, single-center study analyzed electronic medical records, which used hemoglobin A_{1c} (HbA_{1c}) levels as a surrogate for the determination of diabetes status in patients. The study aimed to examine the incidence of new-onset diabetes in patients with gastric cancer who had undergone gastrectomy. Interestingly, the author presents the data via Kaplan-Meier curves, which describe a statistically significant difference, revealing that patients who had an RY reconstruction were more likely to develop new-onset diabetes than patients where surgical reconstruction was achieved via other techniques.

While the findings are interesting, it is essential to enhance the clarity of the study by providing additional information on the sampling methods, the determination of sample size, and a breakdown of the number of events in each group to enable an accurate understanding of study procedures and outcomes. Moreover, an analysis of patients at risk of diabetes before surgery would reduce potential confounding factors. This could be achieved by including a Cox proportional hazard regression to potentially provide more information on the impact of reconstruction methods for the risk of developing diabetes, while also accounting for other covariates. An explanation and

breakdown of other reconstructive techniques (in the OT group) would improve the utility and external validity of this study. Additionally, the participants could have had other comorbidities that could affect the outcome. Therefore, a note on the inclusion criteria and exclusion criteria is necessary.

Below we list major and minor concerns that were discussed by participants of the live review, and where possible, we provide suggestions on how to address those issues.

List of Major Concerns and Feedback

1. There was no rationale provided for the choice between RY and OT. Were any guidelines followed, or was this at the discretion of the attending physician?

Response: The choices were made according to the preferences of the attending physician.

2. Due to the complex nature of postoperative diabetes development, it is crucial to take any confounding variables into consideration and provide a full description of any adjustments made.

Response: I included further detailed demographics (updated in Table 1), and to cope with confounding, I added a propensity score matching analysis.

3. The author should consider including appropriate covariates in the study to assess if they have a confounding effect on the study's result. For instance, is the author able to stratify the patients in terms of their risk of developing diabetes or include relevant information such as family history or concurrent metabolic syndrome?

Response: I added further detailed information in Table 1, but unfortunately, as I have retired from the hospital, information other than what I have collected cannot be implemented.

4. The author should explicitly state the study's inclusion and exclusion criteria. Please consider giving more details on the comorbidities of the included participants. This could be summarized, or tools such as the Charlson Comorbidity Index could be used.

Response: Like in response 3, I was not able to include this information. I am sorry for that.

5. Sufficient details are not provided to allow the reproduction of the study; thus, we suggest you follow the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting. For example, there is content in the Methods section that should go in Results, such as the number of participants included and their baseline characteristics in Table 1. In the same way, information is missing in the Methods section, such as clear definitions of outcomes, statistical analysis, or sample size calculation.

Response: I revised and reshaped the entire manuscript according to the STROBE guidelines.

6. As the cumulative risk of bias for this type of study design is moderate to high, please identify all the variables used in the model. Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Clarify the inclusion and exclusion criteria, together with follow-up time frames and intervals. As the patients underwent surgery between 2005 and 2019, may we assume that the shortest follow-up after surgery was 3 or 4 years?

Response: I added descriptions for these.

7. Also, describe any efforts to address potential sources of bias and explain how the study size was arrived at. Namely, the distribution of the age and sex of the participants is not clear, as there appears to be a bias toward male participants. Refer to the SAGER (Sex and Gender Equity in Research) guidelines for details on conducting a sex-based analysis and disaggregating data according to sex.

Response: The distribution of the age and sex of the participants is clearly stated in Table 1, and I regenerated Table 1 for visuality. Thank you for mentioning the SAGER guidelines. I also added an analysis stratified by sex.

8. Please report the regression model used to assess the associations between the explanatory variables and survival or time to event. How did the author handle learning effects, and the changing and evolving surgical or clinical protocols over the long time frame of this retrospective analysis? Discuss the generalizability of this modeling approach, as well as the direction and magnitude of any potential bias.

Response: The development of surgical techniques during this period is not known. At least no development in surgical technique is known to be involved in the development of postoperative diabetes. No specific direction of bias was assumed, but propensity score matching was performed to address confounding bias.

9. Please include explicit information regarding the competing outcome (ie, mortality events), and justify why no other clinical factors other than HbA_{1c} levels were considered.

Response: Swimmer plots were added to visualize the occurrence of competing outcomes. The onset of diabetes was determined by either the name of the disease in the electronic health record or the HbA_{1c} level.

10. How did the author confirm if the patients were free of diabetes at the time of surgery and before? It would be appropriate if the author provided the baseline (at the time of surgery or before) HbA_{1c} values of the study participants in Table 1.

Response: Since HbA_{1c} is the default test item before surgery, we determined that the patient had diabetes if the HbA_{1c} test or the name of the disease in the electronic medical record mentioned diabetes. We did this by looking at the HbA_{1c} values when we extracted the cases, but we did not state it at the time, and I am sorry that I cannot look back and add it now that I have left the hospital in question.

11. The discussion focused on a procedure that was not mentioned elsewhere or used in this study. Please clarify if this procedure is part of your recommendation for the clinical management of these patients in the future. Additionally, mention if future planned studies will address any stratification of patients for risk of new-onset diabetes mellitus prior to the

surgery, or any analysis of gut microbiota before and after surgery.

Response: The laparoscopic jejunal interposition reconstruction method is relatively new, but we believe that its implementation will increase in the future as no major problems have been identified in previous reports. There are plans to apply for access to the database at the prefectural level and conduct a similar but larger study to this one with a prefectural dataset. If this study is accepted, we intend to ask the database manager to provide the data with the results. Detailed data on gut microbiota, family history of diabetes, and diet are also expected to be included in the dataset that will be submitted for future use.

List of Minor Concerns and Feedback

12. Were any validation techniques used to verify the accuracy of the applied algorithms and analysis, such as code review, unit testing, or cross-validation?

Response: As for code review, it has not been carried out, but the code is public on GitHub, so if something is obviously wrong, it will be pointed out. Also, because it is public, even if imperfections are found in the code in the future, it can be discussed in an open environment.

13. It would be helpful to include a figure explaining the methodology, include more information about the proportion of different reconstructive techniques, and discuss results from other studies to attempt some comparisons for identifying what could have caused similarities or differences in this analysis. For instance, we do not know if the analysis of the groups was blinded.

Response: Groups were not blinded; reviewed information will be included in Table 1.

14. The Methods section lacks proper referencing of previous studies to justify the choice of reconstruction methods (RY vs OT) and the criteria used for defining the onset of diabetes. Referencing previous studies that have investigated similar surgical techniques or criteria for diabetes onset would provide the necessary context and justification for the methods used in the study. Additionally, citing relevant literature would enhance the credibility of the study by demonstrating that the research methodology is grounded in established practices and informed by prior research findings.

Response: The choice of reconstruction method depends solely on the surgeon's preference. For references, similar studies are cited in the *Introduction*—kindly refer to them.

15. Clear visualization of censored data points on a Kaplan-Meier survival curve is essential for accurately interpreting the survival probabilities and understanding the impact of censoring on the analysis. Optionally, you can include CIs for the stratified number of participants.

Response: A Swimmer plot was added.

16. Due to the long time frame of the retrospective analysis and the possibilities of changes in protocol, the author should consider describing how learning effects were handled in the study.

Response: No specific surgical procedures are known to be associated with the development of postoperative diabetes. In addition, the hospital is a training hospital, where surgeons rotate after an average of 2 years and are transferred to other teams (eg, stomach to colon) or to other hospitals, so learning changes are likely to be minimal.

17. There should have been more information about the American Society of Anesthesiologists (ASA) score; this is a subjective score, so even if it was lifted from the electronic record, there ought to be a note pertaining to how many operators assigned the score and the degree of agreement between them.

Response: Discrepancies in ASA scores are indeed a problem, and agreement is said to be around 40% according to the literature. In this paper, the distribution of ASA scores is presented as a reference instead of a comorbidity score. I understand that this should be the Charlson Comorbidity Index or a more diabetes-specific risk and comorbidity assessment, but having left the institution, this is my best-available measure.

18. It is unclear how the missing values were handled. Were they imputed based on a model? What was the definition of an outlier here: greater than 2.5 SDs? What data types are being referred to here? And what inconsistencies needed to be corrected?

Response: In the statistical statement, the treatment of outliers was mentioned, but this statement was deleted because no cases were actually excluded as outliers.

19. What happened to the study participants after 2008 in the OT group (Figure 1)? Why is there a straight line?

Response: The straight line is due to the absence of further diabetes onset in the OT group. A Swimmer plot has been added to make this easier to understand.

20. Please provide more detailed information on what the code does in this study and how it could be used elsewhere.

Response: The entire code has been uploaded to GitHub and is public, so anyone can verify it.

21. In the Abstract, the study setting has been indicated as "Electrical medical records." It should be "Electronic medical records."

Response: Thank you for pointing that out.

Concluding Remarks

We thank the author of the preprint for posting their work openly for feedback. We also thank all participants of the live review call for their time and for engaging in the lively discussion that generated this review.

Response: I changed the colors of lines for visibility. Red (orange) for RY and blue (light blue) for OT.

There is a linear section due to the absence of events, but if this is an obstacle to understanding, it could be replaced by one of shorter duration. The hazard ratios have been calculated for all data, so the various statistics would not change, just that the

Kaplan-Meier curve is easier to understand visually. In Figure 1, both intervals are presented for reference.

References

1. Fairhurst V, Olivier J, Oladoyin O, et al. Peer review of "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study. JMIRx Med 2024;5:e63862. [doi: 10.2196/63862]
2. Onishi T. Incidence of postoperative diabetes mellitus after Roux-en-Y reconstruction for gastric cancer: retrospective single-center cohort study. JMIRx Med 2024;5:e56405. [doi: 10.2196/56405]

Abbreviations

ASA: American Society of Anesthesiologists
HbA_{1c}: hemoglobin A_{1c}
OT: other surgical techniques
RY: Roux-en-Y
SAGER: Sex and Gender Equity in Research
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by E Meinert; submitted 01.07.24; this is a non–peer-reviewed article;accepted 01.07.24; published 14.08.24.

Please cite as:
Onishi T
Author's Response to Peer Reviews of "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study"
JMIRx Med 2024;5:e63859
URL: https://xmed.jmir.org/2024/1/e63859
doi:10.2196/63859

Authors' Response to Peer Reviews of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial"

Sunday Udo[1], MPH, PhD; Pius Ogbu Sunday[1], MSc; Paul Alumbugu Tsaku[1], BSc, MSc, PhD; Israel Olaoluwa Oladejo[1], MPH; Anthony Meka[2], MBBS; Linda Chinonso Ugwu[2], MPH; Motunrayo Ajisola[3], PhD; Joshua Akinyemi[3], PhD; Abiola Oladejo[3], MSc; Akinyinka Omigbodun[3], PhD; Sopna Mannan Choudhury[4], MPH; Jo Sartori[4], MSc; Onaedo Ilozumba[4], PhD; Sam Watson[4], PhD; Richard Lilford[4], PhD

1

2

3

4

Corresponding Author:
Paul Alumbugu Tsaku, BSc, MSc, PhD

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.14.23292603v1

Companion article: https://med.jmirx.org/2024/1/e57310

Companion article: https://med.jmirx.org/2024/1/e56498

Companion article: https://med.jmirx.org/2024/1/e50970

(JMIRx Med 2024;5:e56442) doi:10.2196/56442

KEYWORDS

leprosy; ulcers; wounds; honey; neuropathy; nerves; Africa; randomized controlled trial; RCT

This is the authors' response to peer-review reports for "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial."

Round 1 Review

All the editorial comments are noted and carefully followed.

Anonymous [1]

General Comments

This is an excellent interventional protocol for a randomized controlled trial assessing honey as a potential ulcer therapeutic. Careful consideration has been made to avoid bias and ensure robust results. I would suggest a few things to consider (below) prior to publishing.

Specific Comments

Major Comments

Background and Rationale

- *It is mentioned that 30% to 50% of people infected with leprosy have nerve damage. Be more specific here with "people"—is this a global estimate, American estimate, Nigerian estimate, etc?*

Response: We have now added that it is a global estimate according to a study by Napit et al [2].

- *The background may benefit from a more specific discussion of previous literature. If there is a significant systematic review on the topic—a quick summary of relevant findings in the background (or later in the discussion) would help to situate the rationale behind carrying out such a study.*

Response: We have worked on the literature review; however, we are also mindful of the word limits for this manuscript [3].

Study Setting

- *St. Benedict's Tuberculosis and Rehab Hospital is owned by the Catholic Diocese. Do the authors suspect a potential religious bias in individuals who attend this hospital? Does this affect other social demographics and potentially skew generalizability?*

Response: We do not anticipate any bias associated with the having 1 of the 2 study sites be a Catholic Diocese–owned hospital. This is because the hospital is located in South-South Nigeria, where the citizens are predominantly Catholics. The St. Benedict's hospital is also a well-recognized facility that offers services to people affected by leprosy in the southern region of Nigeria.

Similarly, the Chanchaga Hospital in Niger State, North Central Nigeria, sees most of it's patients from the north. We hoped that having participants from the 2 sites will create some form of balance in our study.

- *I am not sure it is necessary to go into this much detail about the staffing compositions and facilities of each site. Consider truncating.*

Response: We have expunged the unnecessary details about staffing in the 2 facilities.

Additional Consent Provisions

- *The section mentions that the computer program will range-check information. Please specify which computer program.*

Response: The computer program (ie, Research Electronic Data Capture [REDCap]) is now specified in the manuscript.

Intervention Description

- *The honey is being obtained from local bee farmers in North Central Nigeria; however, it is unclear how the honey is being prepared prior to inclusion into the study. I understand it is being checked for botulism (which is great); however—I wonder—is the honey from difference farms being mixed together prior to use? Or is it possible that 1 dressing may be from 1 specific farm, etc? If so, is there a potential risk of interventional procurement bias? Meaning the honey from one farm may be better at wound healing then the honey from another farm? Just something to think about…*

Response: We have a single source for the honey used throughout the study. We try to maintain the integrity of the honey by not processing it other than to filter out the debris.

- *Be more clear about the function of the video recording. Will it also be used to test if assessors can distinguish between honey versus control?*

Response: The video recording at the start of the study is mainly for quality check, including the blinding of the assessors. Once the assessors are happy with the recorded procedure, the clinicians will be asked to proceed with the study.

Confidentiality

- *This section mentions that study forms containing personal identifier information will be kept secured and locked at trial site. Which trial site? Just 1? Or both? Please be more specific here.*

Response: The records are to be maintained at both trial sites. It has now been specified.

Other

- *Will you be collecting demographic data such as sex, gender, creed, socioeconomic status, level of schooling, etc? Would you consider stratifying results by any of these parameters?*

Response: Yes, all the demographic data are included in the baseline data collection on REDCap.

- *Additionally, will you be collecting information on leprosy status, that is, paucibacillary vs multibacillary leprosy or if the patient has progressed to the leprosy reaction stage (type 1 or type 2)? This information may be useful for downstream analysis and can be stratified for to avoid spectrum bias.*

Response: Yes, all the information and more are specified are in the data collection tool.

Minor Comments

Background and Rationale

- *Edit sentence to read: "…and peripheral nerves, causing neuropathy and severe disability, consequently resulting in social exclusion and stigmatization."*

Response: Done.

- *Edit sentence to read: "…new child cases [3], with a grade 2 disability rate of about 15% for the past…"*

Response: Done.

- *Edit sentence to read: "Thirty to fifty percent of…"*

Response: Done.

- *Edit sentence to read: "Ulcers usually occur in anesthetic feet, and will heal slowly with routine therapy, however have a tendency to recur [6]."*

Response: Done.

- *Edit sentence to read: "…documented report record in the Edwin Smith Papyrus…"*

Response: Done.

- *Edit sentence to read: "…gathered and modified by the honeybee…"*

Response: Done.

- *Edit sentence to read: "…exudates, and possesses antimicrobial…"*

Response: Done.

- *Edit sentence to read: "…treatment of difference kinds of wounds, as researchers continue…"*

Response: Done.

- *Edit sentence to read: "…sizeable number of reports that show mixed levels…"*

Response: None.

- *Edit sentence to read: "…with only about 5% of patients reporting pain following dressing."*

Response: Done.

- *The rest of the previous sentence "…and undocumented concern of botulism disease due to infection…" is unclear. Do the same 5% of patients also report concerns of botulism? Or is botulism a concern the authors have, and that has not been reported in previous literature? Either way I would make the botulism argument its own separate sentence that is more clear.*

Response: We have now separated the sentences to make it clearer for the readers to understand.

Study Setting

- *Is it "St. Benedict's TBL" or "St. Benedit's TBL"? Please correct all instances to 1 or the other. The first sentence under study setting uses "St Benedits."*

Response: The correction is done.

- *In "…is a TB and leprosy…" please type out "Tuberculosis" on first use with "(TB)" in quotes as per other abbreviations.*

Response: Done.

Eligibility Criteria

- *Edit sentence to read: "…in the intervention group, all ulcers – not just the one…"*

Response: Done.

- *Edit sentence to read: "Routine swabs will be taken, but the interpretation of…"*

Response: The sentence has been modified in response to a comment by another reviewer.

Additional Consent Provisions

- *Edit sentence to read: "Photograph of the ulcers will be…"*

Response: Done.

Relevant Concomitant Care

- *Edit sentence to read: "…bearing and the level of activity of patients might…"*

Response: Done.

Recruitment

- *Edit sentence to read: "…identified by the on-site clinical…"*

Response: Done.

Plans for Assessment

- *Edit sentence to read: "…database managers at the University of…"*

Response: Done.

Composition of the Data

- *Edit sentence to read: "…Monitoring Committee consists of individuals…"*
- *Edit sentence to read: "…participant has been followed up for 84 days or discharged, whichever…"*

Response: Done.

Dissemination Plans

- *Low- and middle-income countries needs to be fully written out on first use, then the abbreviation "LMIC" can follow.*

Response: Done.

Discussion

- *Edit sentence to read: "…of its near absence from the global health agenda [25], and as such, very little…"*

Response: Done.

- *Edit sentence to read: "A Cochrane review [31] noted that previously published evidence is limited, due to a high or unclear risk of bias (selection, performance, detection, or attrition) detected, imprecision due to little participants, indirectness due to poor outcome measures, and inapplicable interventions."*

Response: Done.

- *Edit sentence to read: "Although honey has been known for centuries to promoted wound healing, there are only a few controlled clinical trials that assess its efficacy."*

Response: Done.

- *This would be a good place to include a brief discussion of relevant findings with specific outcomes or statistics, as I mentioned in the background section.*

Anonymous [4]

General Comments

This paper is a protocol description of an important study, especially for contexts in which advanced wound care products are often not available. It is a well-written protocol with clear steps to take. Below are some of my feedback; I also included some small textual feedback points in the text. You may not be able to address all the points I raised, as it seems that the trial has already started, but in that case, it would be interesting to describe why or why not in the manuscript's text.

Specific Comments

Major Comments

1. Please describe why the 84-day cutoff period was chosen.

Response: The choice of 84 days is based on a recent study by Rai et al [5], which suggests that 80% of leprosy ulcers were healed within 84 days following standard practice. This has been mentioned under the heading "Sample size."

2. The flow chart is a bit small and thus hard to read.

Response: We have attached a full page of the flow chart as a supplementary material to make it easier for readers to understand.

3. Usually, overlapping inclusion and exclusion criteria are not mentioned.

Response: This is noted.

4. It is not clear why hepatitis B or C were added in the exclusion criteria list.

Response: We have considered that hepatitis B and C are not important confounders to the outcome of this study. We have now expunged the statement from the revised protocol.

5. It is not clear for me if patients are clinically admitted or not, and if so, why? For how long? Is this routine care? And what are the discharge criteria?

Response: All the study participants are hospitalized for up to 84 days or discharged when healing occurs before the 84-day period. Those whose ulcers are not healed after the 84-day period will continue with treatment but outside this study depending on the clinicians' advice.

6. If diabetes is excluded, it may be good to also exclude other known reasons for peripheral neuropathy (eg, vitamin B deficiencies).

Response: This is an important suggestion. However, the study is already on course, and we have not considered vitamin B deficiencies from the inception. This is noted for future studies.

7. Are signs of infection also monitored, assessed, or outcome measures?

Response: Yes, the ulcers are beings observed for any sign of infection during the dressing changes. We report any serious adverse event throughout the study.

8. Please explain more on the swabs: what kind of swab is it and what is tested, if this is not part of the research project? In general, it is better to take a routine swab to test for infection (bacterial growth) prior to inclusion instead of prior to randomization, as infection is an exclusion criteria. Also, address this under the heading about "biological specimens."

Response: Thank you for pointing this out. We have now addressed it to clearly show that wound swabs would be taken from the participants to screen for bacterial infections before recruitment. We have also reflected this in the section under "biological specimens."

9. Explain why the video recording is taking place. It may be interesting to also do it with the last 5 patients if it is performed for monitoring reasons.

Response: The video recording is only for the purpose of quality control. The aim is for the independent assessors to verify that the protocol for the wound dressing is carefully followed. We also envisaged that this was only necessary at the start of the study.

10. Why are assessors from Nepal used and not contextual assessors from Nigeria itself (also, is it because of skin color differences of participants in both countries)?

Response: This is part of a multicountry study with India and Nepal. The plan was to send specimens across the countries to be examined by other researchers who understands the study but are completely blinded from the dressing allocation. For each ulcer, the measurement is carried out by 2 separate assessors, in which the measurements are later collated and harmonized by the data monitoring committee. We believe that, by doing so, the risk of bias will be greatly reduced.

11. Please explain why early analysis is taking place after the inclusion of the first three-eighths of participants.

Response: The purpose is to ensure that quality data are recorded throughout the study.

12. I missed the argument in the discussion that mentioned that honey is often relatively cheap and better available than many advanced wound care products.

Response: We have now mentioned that honey is cheaper and a readily available alternative to many wound care products.

13. Include some information about how long data will be stored (number of years), where it will be stored in a secure way, and if it will be shared (pseudonymized) if requested (eg, for reproducibility).

Response: The information is included under "Data management" and "Confidentiality."

Minor Comments

14. Write numbers up to 9 in text.

Response: Noted.

15. Check abbreviations.

Response: All abbreviations checked.

16. Update reference list, include authors, URLs, and "assessed on [date]" in references to websites and online documents.

Response: Noted.

17. Explain the camera used for photography.

Response: The camera used is the Samsung Galaxy Tab S7 (13Mp) as mentioned in the protocol.

18. Please add 2 more references in the discussion.

Response: Noted.

19. Explain more about the pedometer usage.

Response: The pedometers are worn on the nonaffected foot of the participant to monitor their daily step counts. This might show if the level of activity or weight bearing has any impact on the healing rate of the ulcers.

References

1. Anonymous. Peer review of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial". JMIRx Med 2024;5:e56498. [doi: 10.2196/56498]
2. Napit IB, Shrestha D, Bishop J, et al. An individual randomised efficacy trial of autologous blood products, leukocyte and platelet-rich fibrin (L-PRF), to promote ulcer healing in leprosy in Nepal: the TABLE trial protocol. Trials 2021 Jul 15;22(1):453. [doi: 10.1186/s13063-021-05392-5] [Medline: 34266456]
3. Udo S, Ogbu Sunday P, Tsaku PA, et al. Raw, unadulterated African honey for ulcer healing in leprosy: protocol for the Honey Experiment on Leprosy Ulcer (HELP) randomized controlled trial. JMIRx Med 2024;5:e50970. [doi: 10.2196/50970]
4. Anonymous. Peer review of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial". JMIRx Med 2024;5:e57310. [doi: 10.2196/57310]
5. Rai S, Gupta AK, Kumar D, Sharma VP, Agarwal AK. Prospective analytical study of assessment of off loading by total contact cast in treatment of non healing plantar ulcers in anaesthetic foot. Lepr Rev 2016 Mar;87(1):71-77. [Medline: 27255060]

Abbreviations

REDCap: Research Electronic Data Capture

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 01.03.24.

Please cite as:
Udo S, Ogbu Sunday P, Tsaku PA, Oladejo IO, Meka A, Ugwu LC, Ajisola M, Akinyemi J, Oladejo A, Omigbodun A, Choudhury SM, Sartori J, Ilozumba O, Watson S, Lilford R
Authors' Response to Peer Reviews of "Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial"
JMIRx Med 2024;5:e56442
URL: https://xmed.jmir.org/2024/1/e56442
doi:10.2196/56442

Authors' Response to Peer Reviews of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence"

Ramendra Pati Pandey[1,*], PhD; Himanshu[2,3,*], MSc; Gunjan[2,3], MSc; Riya Mukherjee[2,3], MSc; Chung-Ming Chang[3,4,*], PhD

[1]

[2]

[3]

[4]

[*]these authors contributed equally

Corresponding Author:
Chung-Ming Chang, PhD

Related Articles:

Companion article: https://preprints.jmir.org/preprint/51787

Companion article: https://med.jmirx.org/2024/1/e56047

Companion article: https://med.jmirx.org/2024/1/e56440

Companion article: https://med.jmirx.org/2024/1/e51787

(JMIRx Med 2024;5:e55899) doi:10.2196/55899

KEYWORDS

animal-assisted therapy; pet therapy; outcome assessment; policies; systematic study

This is the authors' response to peer-review reports for the paper "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence."

Round 1 Review

Anonymous [1]

General Comments

I enjoyed reading this paper [2]. In general, this is a well-written paper. There are some areas that could be clarified or expanded to improve the strength of the article. Due to the justification, the spacing of punctuation marks appears incorrect, and there are rare instances of double punctuation (periods). The use of American Psychological Association abbreviations at first use was not followed. At times, after providing an abbreviation, the full name is spelled out (eg, "AAT," "PTSD").

Response: We would like to express our heartfelt appreciation for your valuable and pertinent feedback. Your insights have greatly contributed to the improvement of our manuscript. In the following section, you will find our detailed responses addressing each of your comments.

Specific Comments

Major Comments

1. There does not appear to be a Table 2, but it is referenced in the text (page 7).

Response: It seems there may have been some confusion regarding Table 2, which is referenced in the text. I'd like to clarify that Table 2 is indeed included in the manuscript.

2. In the Discussion section on page 15, a reference is made to effect sizes in four outcome areas, yet no effect sizes were reviewed in the article.

Response: Thank you for bringing this to our attention; we have corrected the mistake in the updated manuscript.

3. Also in the Discussion on page 15, the word "power" is used: "The increased number of studies provided greater power in assessing variance heterogeneity and potential group differences." Unless a specific power analysis was performed (if so, it should be discussed), the word "power" could be changed to "support" to reflect a review rather than an analysis.

Response: Thank you for your suggestion. We have replaced the word "power" with "support" as per your suggestion to better align with the sentence.

Similarly, on page 16, the term "meta-analysis" is used. Unless a secondary analysis of pooled data was performed, the term "meta-analysis" should be changed to "analysis." If a secondary pooled analysis was performed, that should be defined and described in the body of the paper.

Response: Thank you for your observation, and I apologize for any confusion. I want to clarify that a secondary pooled analysis was indeed conducted as part of the study, and it has been appropriately defined and described in the body of the paper.

4. Page 16, Society for Healthcare Epidemiology of America is noted as an organization of interest for animal-assisted therapy, as is Pet Partners. The authors may want to consider including the global organization called International Association of Human-Animal Interaction.

Response: Thank you for your insightful suggestion. We included the International Association of Human-Animal Interaction in the updated manuscript.

5. On page 17, the authors cite lack of blinding as a limitation and introduction of bias. I would be curious to know how the authors propose blinding in studies that involve interactions with animals. I strongly suggest this sentence be removed.

Response: Thank you for your suggestion. We have removed the sentence citing the lack of blinding as a limitation and the potential introduction of bias from the updated manuscript

6. The work done by Hinic and others [3] was not a randomized controlled study as noted in Table 1. Please double-check that all studies listed are correctly labeled as randomized studies.

Response: Thank you for bringing this to our attention. We have reviewed the studies listed in Table 1 and have corrected the labeling of the work by Hinic and others [3] to accurately reflect that it was not a randomized controlled study.

Minor Comments

7. Check punctuation for spacing.

Response: Thank you for your suggestion. We have reviewed and corrected punctuation for spacing in the updated manuscript.

8. Check all abbreviations and use abbreviations after first use is defined.

Response: Thank you for your guidance. We have carefully reviewed all abbreviations in the manuscript and ensured that they are defined upon their first use.

9. Check capitalizations in midsentence (page 4: "Dog"; page 7: "Unrepresentative").

Response: We appreciate your attention to detail. Capitalization issues have been addressed and corrected in the updated manuscript.

10. Page 6: "The articles should to be published in English." Wrong tense—change to "were."

Response: Thank you for pointing out the tense issue. We have made the necessary correction in the updated manuscript.

11. Page 1: Three categories of interventions were provided in section 2.4. It would strengthen the paper to include definitions of these categories for the reader.

Response: Thank you for your constructive feedback. We have incorporated definitions for the three categories of interventions mentioned.

Anonymous [4]

Dear Authors,

First of all, your work's topic is up-to-date and meticulously prepared. However, I still have a few questions/suggestions:

Response: We sincerely appreciate and extend our gratitude for your valuable and relevant comments. Your input has been incredibly helpful in enhancing the quality of our manuscript. Below, you will find our point-by-point responses addressing each of your comments.

1. In the Identification section, the total number of articles obtained from each database is given. It is recommended to give separate numbers for each.

Response: Thank you for your suggestion. We have provided the separate number for each database in the updated manuscript.

2. In the box below, the numbers are given as a total, but it may be more appropriate to give separate data for each item.

Response: Thank you for your feedback. We have revised the presentation in the box below to provide separate data for each item.

3. Can keywords be schematized in accordance with PICOS (Population, Intervention, Comparison, Outcomes, Study Design) in the literature review section?

Table...: Keywords used while browsing

Population

Intervention

Comparison

Outcomes

Study design

Response: Thank you for your suggestion. We have now organized the keywords in accordance with PICOS in the literature review section.

4. Has the quality of evidence been evaluated? If so, how was it done? This process can be explained by creating such a subtitle.

- *How did you reduce the risk of bias in studies? Were the articles evaluated and scored separately among authors? Have these scores been analyzed?*
- *By whom and how was the screening done? I think that the most important limitation of this study is that it was scanned by a single person.*

Response: Thank you for your valuable input. We have addressed your concerns and included information on how the quality of evidence was evaluated, the process used to reduce

the risk of bias in studies, and how the screening was conducted in the updated manuscript.

5. In the section where general information is given for the last 16 articles, can it be added which disciplines are studied in particular? Since this subject is studied by various job groups, adding this information can enrich the data. If the mentioned situations are arranged, your article will contribute more to the literature.

Response: Thank you for your suggestion. We have added information about the specific disciplines studied in the last 16 articles in the relevant section.

6. What has been studied in previous systematic reviews? What are the original aspects of this work?

I include below some systematic review studies that may be relevant to the subject:

- *Stern C, Lizarondo L, Carrier J, et al. The experiences and effectiveness of canine-assisted interventions (CAIs) on the health and well-being of older people residing in long-term care: a mixed methods systematic review protocol. PROSPERO. 2020. URL:*https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020161235
- *Whear R, McGill P, Orr N, et al. What are the effects of 'robopets' on the health and wellbeing of older people resident in care homes? A systematic review of qualitative and quantitative evidence. PROSPERO. 2017. URL:*https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017081794
- *Nabi N, McAloney-Kocaman K, Fleming M, Bain S. A systematic review exploring the effectiveness of animal-assisted therapy in improving the psychological well-being of incarcerated individuals. PROSPERO. 2022. URL:*https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022314341

Response: Thank you for your valuable input. We have thought from this aspect too and included the relevant work in the manuscript.

If the mentioned situations are arranged, your article will contribute more to the literature.

I wish you good luck in your work.

Round 2 Review

Response: Thank you for your thoughtful comments and valuable insights regarding our meta-analysis. We appreciate your attention to detail and the opportunity to address the concerns you raised. We have updated the PICOS statement in the updated manuscript.

The following is our rationale for the specific keywords used in our search strategy and clarification of how they relate to the overall objectives of our study:

1. "Pain OR anxiety OR depression OR blood pressure OR BP OR heart rate OR HR": Our focus on these specific health outcomes stems from the recognition that animal-assisted interventions (AAIs) have shown potential effects on psychological well-being (anxiety, depression) as well as physiological parameters (blood pressure, heart rate). Interactions with animals have shown potential in reducing blood pressure and heart rate, which are key indicators of anxiety and stress levels.
2. "Work-related stress OR workplace health OR employee well-being OR burnout" are critical factors that can significantly impact an individual's overall health. Understanding how AAIs influence these aspects helps to recognize the influence of workplace factors on overall health, aligning with a holistic approach to health outcomes assessment.
3. "Tumor OR malignant OR carcinoma OR oncology OR hospitalization OR hospitalized patients OR inpatients" was searched to explore the potential applications of AAIs in health care settings, considering the well-documented positive effects of animal-assisted therapy on patients undergoing medical treatments, including those with cancer. While this may seem at odds with the workplace concepts, our intention was to provide a comprehensive overview of AAIs across various contexts, recognizing their multifaceted applications.

The following are some references pertaining to our search approach linked to the PICOS framework [5-7].

References

1. Anonymous. Peer review of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence".. JMIRx Med 2024;5:e56047.
2. Pandey RP, Himanshu, Gunjan, Mukherjee R, Chang C. The role of animal-assisted therapy in enhancing patients' well-being: systematic study of the qualitative and quantitative evidence. JMIRx Med 2024;5:e51787.
3. Hinic K, Kowalski MO, Holtzman K, Mobus K. The effect of a pet therapy and comparison intervention on anxiety in hospitalized children. J Pediatr Nurs ;46:55-61. [doi: 10.1016/j.pedn.2019.03.003] [Medline: 30852256]
4. Anonymous. Peer review of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence". JMIRx Med 2024;5:e56440.
5. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7:16. [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
6. Eriksen MB, Frandsen TF. The impact of Patient, Intervention, Comparison, Outcome (PICO) as a search strategy tool on literature search quality: a systematic review. J Med Libr Assoc 2018 Oct;106(4):420-431. [doi: 10.5195/jmla.2018.345] [Medline: 30271283]

7. Evidence based medicine: the PICO framework. Icahn School of Medicine at Mount Sinai. URL: https://libguides.mssm.edu/ebm/ebp_pico [accessed 2024-03-07]

Abbreviations

AAI: animal-assisted intervention
PICOS: Population, Intervention, Comparison, Outcomes, Study Design

Edited by E Meinert; submitted 28.12.23; this is a non–peer-reviewed article;accepted 28.12.23; published 18.03.24.

Please cite as:
Pandey RP, Himanshu , Gunjan , Mukherjee R, Chang CM
Authors' Response to Peer Reviews of "The Role of Animal-Assisted Therapy in Enhancing Patients' Well-Being: Systematic Study of the Qualitative and Quantitative Evidence"
JMIRx Med 2024;5:e55899
URL: https://xmed.jmir.org/2024/1/e55899
doi:10.2196/55899

Author's Reponse to Peer Reviews of "Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers"

David Bychkov, PhD

UC Institute for Prediction Technology, University of California, Irvine, Orange, CA, United States

Corresponding Author:
David Bychkov, PhD

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.08.14.23294073v1

Companion article: https://med.jmirx.org/2024/1/e57701

Companion article: https://med.jmirx.org/2024/1/e57159

Companion article: https://med.jmirx.org/2024/1/e52198

(JMIRx Med 2024;5:e57116) doi:10.2196/57116

KEYWORDS

TRICARE; health care fraud; Defense Health Agency; fraud; fraudulent; insurance; coverage; beneficiary; beneficiaries; background check; background checks; demographic; security clearance; FDA; Medicaid; Medicare; provider; provider referral; military; False Claims Act; HIPAA breach; OIG-LEIE; inspector general; misconduct; insider threat; information system; zero trust; data management; Food and Drug Administration; Health Insurance Portability and Accountability Act breach; Office of Inspector General's List of Excluded Individuals and Entities

This is the authors' response to peer-review reports for "Insider Threats to the Military Health System: A Systematic Background Check onf TRICARE West Providers."

Round 1 Review

Anonymous [1]

General Comments

In general, the manuscript [2] is informative and includes a lot of information on health care providers who participate in TRICARE insurance. The study examines those who have received some sort of exclusion, sanction, or other reprimand based on health care fraud or harm. This study is timely and has practical implications for protecting patient care, particularly for those who are in a vulnerable position such as veterans or warfighters. I hope the following comments are taken as constructive criticism and interest in the overall improvement of the study. I appreciate the opportunity to review this study.

Below is a list of important fixes that I recommend considerable time be spent on and some minor fixes. In general, I think the key limitation of the study is that it can better state the significant contribution of the study. I understand the need for such a study, but as it stands, the study can further improve by spending more time on why and how health care providers land on exclusion lists such as the Office of Inspector General's (OIG) List of Excluded Individuals and Entities (LEIE). Indeed, the study uses several databases that exclude physicians or provide reasons why a physician no longer participates in such programs, but the author can improve their justification for the study on why this is needed.

The second key limitation of the study is the Methods and Results section. In particular, this section needs improvement with clearer detail and justification on why the author had a selection criterion (vs examining all zip codes). In addition, the Results section can improve with better organization of the findings. As it reads, the results are a bit difficult to follow with all the zip codes laid out.

Last, the study could benefit from greater discussion on the implications of the study. At the moment, it pushes for more transparency, but the author could use their data more to discuss the impact of their findings. For example, why would publishing the National Provider Identification (NPI) numbers help patients? What do patients or the author want to gain from that transparency? How can this help future patients or hold physicians more accountable? The discussion loosely taps into the implications, but the study could really tease out this argument more.

Overall, the study was easy to follow and did provide some interesting content to consider. I think the study can better serve

the public and has great implications! I would like to see these implications highlighted more so that the reader can really see the contribution the study makes.

Specific Comments

Major Comments

1. Introduction: Provide an explanation of what the OIG LEIE is for the reader. It is important to inform the reader that the OIG LEIE excludes participation in the program for various reasons—not just a quality-of-care issue. For example, the OIG LEIE also can exclude physicians on a financial offense matter. This helps the reader understand the gravity of the situation, particularly when the author discusses the increased risk of mortality and hospitalization of these patients. In short, I would like to see more development and background on how individuals are included in the LEIE to increase awareness for the reader.

a. For more information about LEIE and how physicians are placed on the list, please see the following: Burton B, Sun D, Jesilow P, Pontell HN. Two paths, one destination: a demographic portrait of physicians sanctioned by the federal government. J Health Hum Services Adm. 2022;45(3):142-180.

Response: I have tried to add additional context about all the various lists, including the OIG LEIE. Ironically, I suspect we learned the most about TRICARE administrators from my article. Where they do enforce, there are exclusion-provider name matches.

2. Methods: This section needs improvement. First, please provide more justifications and in-text citations to justify the methods used for the study. This will help strengthen the Methods section. As it reads now, there seem to be no prior studies listed that use this method (although that is not the case). Second, why did the author limit the search to the "83 most populous zip codes"? Why not include all zip codes? Does this relate to the number of people participating in TRICARE, or is this because there are simply more people living in those zip codes? Please include a justification here on why there is a population cutoff. Third, on page 8, the author writes that there were 22 states that were included, but the list only included 21 states (from my understanding). In addition, why were some of the zip codes (eg, St. Louis, Rock Island Arsenal) excluded and others included (eg, Amarillo, Lubbock, and El Paso areas only for Texas)?

Response: I have done my best to expand on the Methods section. I went back and had to comb through all the raw data. In fact, a large number of zip codes contain no humans (just raw land or no population). As western states are less population dense than TRICARE East states, the easiest solution was to use a minimum population as a filter. Unfortunately, I have to run searches manually (rather than via Python) to comply with government regulations. Otherwise, yes, I could just hack TRICARE West and search everything in 5 minutes! The project took over 1 year and resulted in a whistleblower complaint. You can imagine the nightmare.

22 versus 21 states: TRICARE West covers only 21 states. For whatever reason, they decided to include zip codes and a state (containing no providers) from non–TRICARE West regions. To simplify things, I excluded those meaningless data from this revision.

3. Results: The author generally states that their findings are consistent with past research but do not include a list of articles to which they are referring. Only one article is referenced [3]. Please provide further support for that claim (in other words, please include all other studies to support the claim of consistent findings). In addition, when discussing results like on page 10, the presentation is difficult to follow with all the zip codes listed and separated by a hyphen. Please consider reorganizing this presentation or placing the list of zip codes in a footnote to ease the presentation of results.

Response: I added more data on consistency

I added a whole section on significance.

4. Discussion: The significance of the study could be further elaborated on. At the moment, it pushes for more transparency, but the author could use their data more to discuss the impact of their findings. For example, why would publishing the NPI numbers help patients? What do patients or the author want to gain from that transparency? How can this help future patients or hold physicians more accountable? The discussion loosely taps into the implications, but the study could really tease out this argument more.

Response: I tried to clarify how/why NPI numbers could help patients. I also tried to clarify the limitations that administrators face as a result of health care labor shortages.

Minor Comments

1. Clean up the grammar and punctuation. For example, on page 4, the author states, "Nicholas et al performed a cross-sectional study of 8204 Medicare beneficiaries who received care from excluded providers. It revealed that patients treated by fraudsters experience a 13%-23% increased risk of mortality and 11%-30% higher risk of hospitalization (Nicholas et al, 2019)." Note, that the start of the sentence, "Nicholas et al" needs a period and a year in the citation.

2. I suggest adding a numerical list when discussing the different databases that are available for searching a physician. For example, on page 5, the author lists several different databases starting with the sentence "Multiple public databases exist to search names with respect to each of these issues, including..." Adding in a numbered list can make the information more digestible for the audience. This can also be cleaned up (ie, adding a numeric list) on page 7 when listing the different databases that the physicians were screened in.

3. Page 6, it is stated that 203 names appeared in up to 3 additional types of databases. However, what are these 3 additional types of databases? Is it referring to the earlier-mentioned databases? This is unclear.

Response: I fixed the grammar and other issues.

Anonymous [4]

General Comments

This paper examines the list of TRICARE providers eligible to deliver telehealth whose names appear on one or more federal sanction lists. This work could have a high impact with implications for national security and patient safety. However, it is not well organized and does not seem to adhere to the scientific method.

Specific Comments

Major Comments

1. This is important work, but is it actually science? A team compared two lists. There are no statistics, minimal numbers, and only one hard conclusion (improper monies). Lots of speculation, but no real answers. Because you are calling out the Defense Health Agency (DHA) and the Military Health System (MHS) on inadequate oversight, your conclusions must be driven by airtight methodology and presented in a professional and well-organized manner. Otherwise, you may just submit this work to the DHA's OIG as you've already done and call it completed.

Response: I reorganized the content and attempted to obey the scientific method.

2. Assuming you decide to go with the science, this article has important things to say, but it is not yet ready for publication. It is poorly organized, somewhat informal in tone, and comes across as inflammatory in places. An example of poor organization is the focus on cyber threats and potential in the Introduction, improper monies paid to sanctioned providers in the Results, and a distrust of provider data in the Discussion. Patient safety is not discussed until page 17. Recommend mentioning all these issues in the Introduction and then addressing them in the Results/Discussion in a systematic, organized fashion. Also, streamline areas where the same data is repeated multiple times.

3. Similarly, I recommend keeping all the DHA/MHS recommendations together and at the end of the Discussion section, and addressing these in a systematic and organized fashion. "Based on these results, the DHA/MHS/ TRICARE/ whoever should consider the following: (1) Recommendation 1. (2) Recommendation 2..." etc. (Don't need to take my wording but this is a general idea.)

Response: I attempted to reduce the inflammatory language, place patient safety stakes up front, and group the core issues in the Introduction.

I also organized the data in a cleaner way using Google Maps to emphasize geographic relationships between the data sets.

Finally, I provided substantially more evidence to support my correlations and fewer opinions.

4. If there are other people who participated in this study, they should be included as authors or in the Acknowledgments. I doubt that one person compared tens of thousands of names solo.

Response: Unfortunately, I am a military employee with no grant funding. I conducted all of this work alone over 12 months. I did my best to acknowledge my colleagues, but I had no assistants or other helpers.

References

1. Anonymous. Peer review of "Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers". JMIRx Med 2024;5:e57701. [doi: 10.2196/57701]
2. Bychkov D. Insider threats to the military health system: a systematic background check of TRICARE West providers. JMIRx Med 2024;5:e52198. [doi: 10.2196/52198]
3. Chen A, Blumenthal DM, Jena AB. Characteristics of physicians excluded from US Medicare and state public insurance programs for fraud, health crimes, or unlawful prescribing of controlled substances. JAMA Netw Open 2018 Dec 7;1(8):e185805. [doi: 10.1001/jamanetworkopen.2018.5805] [Medline: 30646294]
4. Anonymous. Peer review of "Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers". JMIRx Med 2024;5:e57159. [doi: 10.2196/57159]

Abbreviations

DHA: Defense Health Agency
LEIE: List of Excluded Individuals and Entities
MHS: Military Health System
NPI: National Provider Identification
OIG: Office of Inspector General

Edited by E Meinert; submitted 05.02.24; this is a non–peer-reviewed article;accepted 05.02.24; published 09.04.24.

Please cite as:
Bychkov D
Author's Reponse to Peer Reviews of "Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers"
JMIRx Med 2024;5:e57116
URL: https://xmed.jmir.org/2024/1/e57116
doi:10.2196/57116

Authors' Response to Peer Reviews of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"

Usha Lokala[1], MS; Orchid Chetia Phukan[2], MTech; Triyasha Ghosh Dastidar[3], MSc; Francois Lamy[4], PhD; Raminta Daniulaityte[5], PhD; Amit Sheth[1], PhD

1

2

3

4

5

Corresponding Author:
Usha Lokala, MS

Related Articles:

Companion article: https://arxiv.org/abs/2304.10512

Companion article: https://med.jmirx.org/2024/1/e58317

Companion article: https://med.jmirx.org/2024/1/e58320

Companion article: https://med.jmirx.org/2024/1/e58321

Companion article: https://med.jmirx.org/2024/1/e48519

(JMIRx Med 2024;5:e57838) doi:10.2196/57838

KEYWORDS

opioid; substance use; substance use disorder; social media; US; opioid crisis; mental health; substance misuse; crypto; dark web; users; user perception; fentanyl; synthetic opioids; United States

This is the authors' response to peer-review reports for "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study."

Round 1 Review

Responses to Reviewers' Comments

The below includes the comments from reviewers and the authors' responses to each comment. We thank all the reviewers for their suggestions and comments.

Reviewer O [1]

1. The study [2] acknowledges the challenges of crawling crypto markets and the restricted crawling process, which limits the data available for analysis.

Need to explain in the manuscript.

Response: The manuscript is revised as per this comment. Crawling crypto markets poses a significant challenge when applied to data science and machine learning to study the opioid epidemic due to the restricted crawling process. However, we made our proposed data set available to the research community for further analysis.

2. The study proposes building an opioid drug social media knowledge graph (ODSM-KG) but does not provide details on the potential impact or implications of such a graph.

Need to provide the details in the manuscript.

Response: The manuscript is revised as per this comment. The impact of the ODSM-KG is described in the manuscript as follows: "To identify the approaches for mitigating the misuse of opioids, it is imperative to study consumption patterns at the national and regional levels, the influences of the pharmaceutical industry, and the sociopolitical determinants that affect consumption. Our ODSM-KG aims to create a web-based tool that will allow for the depiction of historical patterns and enable comparisons between opioids, time periods, and areas within the United States. Given that our ODSM-KG was primarily based on data, we aim to enhance its accuracy and effectiveness by seeking guidance from a subject matter domain expert. This will enable us to customize it for unique scenarios and cater to the needs of specific users."

3. The study explicitly states that it does not make any clinical diagnosis or treatment suggestions, which indicates a gap in translating the research findings into practical applications for addressing substance use disorder (SUD).

Need to justify how this study will be helpful for clinical situations.

Response: The manuscript is revised as per this comment. While the study may not directly provide clinical diagnosis or treatment suggestions, its findings can still be highly valuable for informing clinical practice and interventions for SUD in the identification of risk factors. Even if the study does not prescribe specific treatments, its findings can inform the development of novel intervention strategies or the optimization of existing ones. For example, as our paper highlights certain sentiment and emotion patterns associated with SUD, clinicians can incorporate this knowledge into their approaches. This research also contributes to understanding the factors contributing to the onset of SUD, which can help in designing prevention programs aimed at at-risk populations, such as individuals with mental health disorders. Research findings can inform policy makers about the effectiveness of current strategies and the need for adjustments in regulations, health care policies, or resource allocation for addressing SUD at a broader level.

Anonymous [3]

1. The paper is well written and easy to understand. See comments below for a summary description of the paper from my perspective.

2. However, I would have liked to see insights ideally established in the medical literature and supported by the experimental context in this paper (eg, those that can substantiate the prediction results and how this type of artificial intelligence can benefit SUD-related outcomes).

Response: We provided a comprehensive review of relevant studies and findings from the literature that support our approach and the predictive results obtained through our artificial intelligence methodology. The experimental context is supported by the literature described in our related work. This also supports our insights and prediction results as per the literature. We discussed the study's potential impact on how artificial intelligence can benefit SUD prediction in the Discussion section, the Global Relevance section, and the Results section.

3. Although a temporal pattern–aware method is implemented in this paper, which is a big positive, I would like to see an analysis over two distinctly separate time periods to establish the consistency and robustness of the proposed approach.

Response: Figures 4-6 show the consistency of sentiments, topics, and emotions for drugs since 2015. The time period considered is between 2015 and 2020 with a side-by-side comparison of drugs.

Anonymous [4]

This study can be considered for publication if the researchers are able to revise it to improve the clarity of the objective, theoretical relevance, and practical value. I will suggest that there should be a segment on the objective of the study immediately after the introduction. This will help in giving the study a direction. Please, include this citation:

Obosi AC, Fatunbi AM, Oyinloye O. Peer pressure and substance use as predictors of mental health among in-school adolescents in Nigeria. Ianna J Interdisciplinary Stud. 2022;4(1):1 - 9.

Explain the implication of your findings to other countries, that is, give the study an international outlook.

Response: This paper by Obosi et al is mentioned and cited in the related work sections of the paper. Thank you for your suggestion. A section named "Global Relevance" has been added to the paper now as per your comment, and the impact is described.

References

1. Subramaniyan V. Peer review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study". JMIRx Med 2024;5:e58317. [doi: 10.2196/58317]
2. Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A. Detecting substance use disorder using social media data and the dark web: time- and knowledge-aware study. JMIRx Med 2024;5:e48519. [doi: 10.2196/48519]
3. Anonymous. Peer review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study". JMIRx Med 2024;5:e58321. [doi: 10.2196/58321]
4. Anonymous. Peer review of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study". JMIRx Med 2024;5:e58320. [doi: 10.2196/58320]

Abbreviations

ODSM-KG: opioid drug social media knowledge graph
SUD: substance use disorder

Edited by E Meinert; submitted 27.02.24; this is a non–peer-reviewed article;accepted 27.02.24; published 01.05.24.

Please cite as:
Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A
Authors' Response to Peer Reviews of "Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study"
JMIRx Med 2024;5:e57838
URL: https://xmed.jmir.org/2024/1/e57838
doi:10.2196/57838

Authors' Response to Peer Reviews of "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study"

Silvia Corbera[*], PhD; Amanda M Marín-Chollom[*], PhD

Department of Psychological Science, Central Connecticut State University, New Britain, CT, United States

[*]all authors contributed equally

Corresponding Author:
Silvia Corbera, PhD

Related Articles:

http://preprints.jmir.org/preprint/52970

https://med.jmirx.org/2024/1/e59430

https://med.jmirx.org/2024/1/e52970

(JMIRx Med 2024;5:e58859) doi:10.2196/58859

KEYWORDS

prosocial behavior; COVID-19; anxiety; COVID-19 prevention; preventive health behavior; adherence to prevention

This is the authors' response to peer-review reports for "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study."

Round 1 Review

Summary Provided by Reviewers [1]

The study [2] investigates the complex interplay between anxiety (both state and trait), prosocial behaviors, and adherence to COVID-19 preventive measures among college students. While overall prosocial behaviors did not directly correlate with anxiety, a seemingly significant crossover effect emerged in relation to public prosocial behaviors, suggesting that individuals with lower self-oriented tendencies exhibited increased adherence behaviors under heightened state anxiety. The study used a quantitative research design with a sample of 54 undergraduate students, using online questionnaires to measure various psychological factors and preventive behaviors. Individuals with high anxiety showed increased adherence to preventive measures, contrary to the hypothesized moderating effect of prosocial behaviors. Overall, the reviewers appreciated the effort and recognized the challenges of conducting a research study in the context of an unprecedented social condition. However, the findings are challenged by weak effect sizes, multiple comparisons, and unclear appropriateness of using prosocial behaviors as a moderating variable. The study underscores the psychological impact of the COVID-19 pandemic on college students and suggests the need for further exploration into the nuanced relationships between anxiety, prosocial behaviors, and adherence to public health guidelines. Despite its strengths in data collection and questionnaire use, limitations such as a narrow participant pool and reliance on self-reporting warrant cautious interpretation of the results. The study encourages future research to delve deeper into these intricate connections, offering insights into potential interventions for promoting adherence to COVID-19 preventive measures and beyond.

Response: We thank the editor and reviewers for their valuable time and insightful and helpful comments. In this resubmission, we have provided answers to your comments below and improved the manuscript following your contributions.

Below we list major and minor concerns that were discussed by participants of the live review, and where possible, we provide suggestions on how to address those issues.

List of Major Concerns and Feedback

Small Sample Size and Mediation Analysis

One of the main concerns raised in the discussion was the small number of study participants. This is acknowledged as a limitation factor in the discussion, but what is less clear is if a mediation analysis is the right approach to analyze these data. One reviewer suggested the use of a multivariate analysis instead as it would take all variables into account without forcing potentially artificially generated causal mediations between variables that don't show an obvious causal dependency. Another reviewer, however, felt that while this approach may work and it would be useful to explore, using a

multivariate analysis may lead to overfitting in most covariates given the small sample size and given that 86% of subjects reported anxiety. Overall, the suggestion is for the authors to provide a rationale for selecting anxiety as a mediator variable over prosocial tendencies, or vice versa, and possibly explore other analyses and comment on the limitations of the approaches.

Response: We thank the reviewers for this comment. We wanted to clarify that this study used a moderation analysis, not a mediation analysis. The goal was not to examine causality but the moderating effects of prosocial tendencies in the relationship between anxiety and adherence to preventive behaviors to prevent the spread of COVID-19. Given that the study was conducted during the second wave of the pandemic, the reported stress levels of the undergraduate students who participated in the study were high (86% of subjects reported anxiety); the goal of the study was to examine the relationship between stress levels and adherence to preventive behaviors to prevent the spread of COVID-19, and if this relationship was moderated by prosocial tendencies. Based on the scientific literature, which robustly indicates that stress influences decision-making processes [3,4], we wanted to examine if this finding would be replicated instead of using laboratory stress-inducing techniques, using the COVID-19 pandemic as the natural worldwide stressor and the decision-making process of adhering to preventive behaviors. However, the scientific literature is scarce in regard to the role of prosocial behaviors in situations of stress, and therefore, we were particularly interested in examining the role of prosocial behaviors in individuals who would show high versus low prosocial behaviors.

In regard to the small sample size, we performed a power analysis using SPSS with a power (β) set at 0.8, the Cohen *f* at 0.15 for medium effects, and significance level at .05. The sample needed was 55. Therefore, we consider that the study was sufficiently powered as our study had 54 participants. We acknowledge that the sample size was small, but the study was completed during a second wave of COVID-19, with a limited number of participants at the time. We were particularly interested in examining the role of anxiety at that moment, given the exceptionality of the situation. Worldwide life circumstances changed greatly after that spring (with full availability of the vaccines to everyone including young adults) and have changed now, and the study's historical environment is different from what we captured, so it was not possible for us to add more participants to this study. We have now acknowledged this further in the Limitation section of the manuscript.

Uniform/Convenience Sampling

The reviewers acknowledged the study's challenging circumstances and the effort to capture unique data. However, they expressed concern about generalizability, noting that all participants were undergraduates from one college. The demographic homogeneity of this group may limit applicability to diverse populations, raising caution about extrapolating results across age groups, educational backgrounds, and cultural contexts. The convenience sampling method may have introduced bias, as easily accessible participants might not represent the broader target population.

Response: We thank the reviewers for this insightful comment. We acknowledged in the Limitation section that the sampling method used was convenience sampling with an undergraduate college population. We agree with the reviewers that this homogeneous sample may limit the applicability to diverse populations, and therefore, we now further clarified this limitation as suggested by the reviewers by raising caution about extrapolating results across age groups, educational backgrounds, and cultural contexts. At the same time, we wanted to point out the idiosyncratic developmental challenges of the undergraduate population, as for most of them, the college experience is already a significant change in their life, let alone in the middle of a worldwide pandemic. Research has shown that the undergraduate population was a specific vulnerable population affected greatly by the stressors of the COVID-19 pandemic, especially among the lower socioeconomic undergraduate groups [5,6]. Therefore, even though the sample was not representative of other age groups, we consider that it merits study. We have now further clarified this item in the Limitation section.

University Policy and Mask Mandates

The study doesn't explicitly consider the potential impact of university campus policies during the pandemic, such as mask mandates and social distancing, on adherence behaviors. The findings may be influenced by the specific characteristics of the chosen university, including its restriction policies. For example, students on campus may wear face masks due to safety requirements when entering campus common spaces versus through their own personal decision process. Therefore, exercising caution in generalizing conclusions to broader contexts is suggested. Reviewers highlight the importance of future research with more diverse samples to enhance external validity, reinforcing the study's overall robustness. This provides a constructive pathway for refining the study's scope and applicability.

Response: We thank the reviewers for this comment. We stated in response to the "Uniform/Convenience Sampling" comment that we added further information in the Limitation section regarding the caution of extrapolating the results to other populations and contexts. We agree that the university policies at the time were exceptional. However, we wanted to note that the measure that we used to examine the frequency of engaging in COVID-19 preventive behaviors asked questions that were aimed at all aspects of their life not specific to any context, and in addition, they also had the option to answer "I don't know/prefer not to answer/Not applicable" for each question. Therefore, participants had the option of not answering and not answering because the question was not applicable to their life such as "Avoiding Playdates (letting children play with other children)." Specifically, the questionnaire asked "Please indicate the frequency with which you have adopted each action/behavior in the previous 7 days:" in which they had to answer items such as "Hand washing with soap and water," "Using a hand sanitizer," "Wearing a Face Mask," "Avoiding non-essential travel," and "Avoiding opening the mail or delivered goods" using a 5-point scale ("Most of the time," "Some of the time," "Seldom," "Never," and "I don't know/prefer not to answer/Not applicable"). At the time of the pandemic in the spring of 2021,

the safety measures applied by the university were many, including social distancing and mask wearing, but also followed state-mandated measures as the state had a travel advisory and a requirement to complete a travel health form and quarantine for a 10-day period if traveling outside of Connecticut for more than 24 hours in states other than New York, New Jersey, and Rhode Island or countries other than the United States. Therefore, this mandate was statewide. Additionally, at the time, spring 2021, vaccinations were provided for priority populations such as those older than 75 years, and vaccinations for those 65 years and older were not planned until mid-February. Vaccinations for everyone 16 years and older were not available until April 1. Additionally, there was a ban for in-person research during the spring of 2021 that was not expected to be lifted until May 2021, which is when the semester was over. We believe that the situation was exceptional on campus and outside campus, and the measures on campus were designed for safety. Additionally, students had the option of not answering the items of the questionnaire; therefore, we believe that the likelihood that the university policies influenced their responses to be biased in any way was very small.

The study's cross-sectional design and reliance on self-reporting introduce potential limitations in establishing causal relationships and accurate data collection. (In general, no retrospective exploratory study can show causality, asserting a causal relationship amounts to the post hoc ergo propter hoc logical fallacy.) The one-time nature of the study also limits insights into the dynamic nature of psychological factors and preventive behaviors over time. Caution should be exercised in interpreting the results, as correlation does not imply causation.

Response: We thank the reviewers for this comment, and we added this in the Limitation section. In addition, in our Discussion section, when interpreting the results, we did not imply causation in any instance, as we were carefully using the words "association" in lines 1 and 2 of the discussion: "The present study aimed to examine the association between state and trait anxiety during the COVID-19 pandemic and the adherence behaviors to prevent the spread of COVID-19..."; we also use the phrase "more likely to" such as in line 6 of the Discussion: "...we hypothesized that participants with high state and trait anxiety would be more likely to adhere..."; also, the word "relationship" was used in line 10: "...and that this relationship would be stronger for individuals with high prosocial behavior tendencies...." Therefore, we were careful to not imply causation. We changed the wording in line 33 of the Discussion to ensure causality was not implied.

Furthermore, reviewers don't think that the results of this study can be used on their own to make any definitive public health policy recommendations.

Response: We agree and we acknowledge this in the Limitation section but can inform other future studies on the role of prosocial behaviors in the relationship between stress and adherence to preventive behaviors. We added the following statement: "Fourthly, because of the small sample size, the cross-sectional nature of the study, this results on their own cannot make any definitive public health policy recommendations but can inform other studies on the role of prosocial tendencies in the relationship between stress and adherence to preventive behaviors."

Adherence Scores

The study mentions adherence to COVID-19 public health safety recommendations as an outcome variable. There is a need for more clarity on how the adherence scores were calculated, especially considering potential confounding factors such as university campus policies during the pandemic.

Response: We thank the reviewers for these comments. We provided information regarding the influence of "campus policies during the pandemic" in the answer to the "University Policy and Mask Mandates" comment. Also, regarding the calculation of the adherence scores to COVID-19 preventive behaviors, we described the calculation in the Methods section. Please see below the wording from the Methods section:

"COVID-19 Preventative Behaviors. The COVID-19 International Survey (CIS) from the PhenX Toolkit (2020) was used to collect data on what preventative COVID-19 behaviors participants were engaging in. The 23 items within the survey were specific to what frequency individuals are engaging in COVID-19 preventative behaviors were used for this study (eg, hand washing, mask wearing physical distancing, avoiding social gatherings, self-quarantining after travel, self-quarantining if infected or likely infected were used). These items are answered on a 4-point Likert scale from never to most of the time, with the additional option of 'don't know/I prefer not to answer/Not applicable.' One total sum score was calculated, and higher scores indicated higher engagement in COVID-19 preventative behaviors."

Missing Data

The study's approach to handling missing data in nonmandatory survey questions is not explicitly discussed. This may impact the results and should be clarified.

Response: We thank the reviewers for this insightful comment. Most participants had complete data for the variables used in the analysis, except for age (n=53) and trait anxiety (n=44). State anxiety had complete data (N=54). The lack of effects with trait anxiety could be due to the smaller sample size, and state anxiety may be more relevant to the COVID-19 pandemic context. We have now added this information in the Discussion section.

Reliability Metrics

The study does not provide information on test-retest reliability, accuracy against a gold standard, or error of measurement for the Prosocial Tendencies Measure (PTM) reliability. Reliability induction from other studies is mentioned, but the study population's specific reliability is not demonstrated. Without these critical reliability metrics, the study leaves a gap in the assessment of the psychometric properties of the PTM. Including such information would enhance the transparency and credibility of the study's findings, allowing readers to better evaluate the reliability and validity of the instrument used to assess prosocial behaviors. Future research may consider providing a comprehensive assessment of the psychometric

properties of measurement instruments to strengthen the methodological rigor and overall quality of the study.

Response: We thank the reviewers for this comment. We have now added the validity and reliability information of this measure in the Methods section of the manuscript.

Statistical Model and Data Selection

Some reviewers expressed concern related to the lack of transparency about how variables were selected as moderators or mediators, how some others (eg, age) were chosen to be excluded, and how others were chosen to be reported on from the cited "larger study." Adding clarity around the rationale that led to making such choices would help the reader better contextualize the results.

Response: We thank the reviewers for this comment. In this study, we used a moderation analysis, not a mediator analysis. The variables were chosen based on the literature and theory that drove the research protocol. In our institutional review board (IRB) protocol (#20089), we proposed testing two models, and these results are from one of those models driven by the literature.

Furthermore, a scoring guide for the CIS Survey would be helpful to add. There is a concern that a simple sum method may be biased because some questions may not be relevant to all subjects (eg, playdates only impact subjects that have childcare responsibilities).

Response: We thank the reviewers for this suggestion. We provided information on this scale in the Methods section and in response to the "University Policy and Mask Mandates" and "Adherence Scores" comments. We want to emphasize that participants had the option to answer "I don't know/prefer not to answer/Not applicable" for each question. Therefore, participants had the option of not answering if it was not applicable to them, such as "Avoiding Playdates (letting children play with other children)."

Ethics

While the study mentions obtaining IRB approval and online passive consent, specific details regarding confidentiality, privacy safeguards, and participant understanding of risks are not thoroughly addressed.

Response: We agree and we disclosed that this study was fully IRB approved. At Central Connecticut State University, studies that are IRB approved are shown to follow proper confidentiality procedures and privacy as well. All these procedures were followed in this study, as per the IRB approval obtained in this IRB protocol (#20089).

Furthermore, it is not clear what the authors mean by "passive consent."

Response: We agree and we understand the confusion, and we removed the word passive consent and added the following wording: "Participants were presented with an online informed consent, and they acknowledged it by pressing a button to continue the study."

Data and Reproducibility

The study provides a moderate level of detail, but more specific information is needed for reproducibility. This includes additional demographic details, exact questionnaire wording, and more details on moderation analyses. The study would benefit from providing a more comprehensive set of demographic information about the participants such as age distribution, gender distribution, and other relevant characteristics. A richer demographic profile would contribute to a more nuanced understanding of the study population and facilitate comparisons with other research. Reviewers suggested adding available details to Table 1.

Response: We thank the reviewers for this insightful comment. We now added additional demographic details in the manuscript describing the participants further. As described now in the manuscript, we added the age range and percentiles, and we added a full table—a new Table 1 that includes a description of the demographic variables: gender, race/ethnicity, enrollment status, whether they are first generation students, marital status, employment status, hours of work per week, and housing situation and living situation.

While the study mentions that data are available upon reasonable request, reviewers suggest considering providing additional information on how interested researchers can request the data, perhaps from the corresponding author or another designated contact. This could enhance transparency and facilitate potential collaborations or further scrutiny of the results.

Response: We agree with the reviewers on this comment, and we now added in the Data Availability statement that "the data generated from this study can be available from the corresponding author upon reasonable request."

List of Minor Concerns and Feedback

Readability

Overall, the reviewers thought that the manuscript would benefit from a clearer explanation of key terms and recommended keeping the terminology consistent across the manuscript so as to help the reader better follow the narrative and interpret the findings. For example, there was some confusion among reviewers on the meaning of "public prosocial scale."

Response: We thank the reviewers for this comment, and we now have added extensive clarifications in the Discussion section to enhance the readability of the manuscript and concretely remove any possible confusion with the definition of the public prosocial scale.

Approach and Results

It may be helpful to show more information about some of the background variables. One question is if the deviation of age from a normal distribution is significant and thus a possible contributor to the study's findings if age correlates with adherence or anxiety. Showing not only mean and SD but also median, quartiles, and range may provide a better feel for what the study population, or at least the participant sample, is like.

Response: We thank the reviewers for this comment. We had already included the mean and SD of age, and we now added the range and the quartiles. Additionally, we now provide an extensive list of demographic variables in Table 1. As the reviewers can see, the sample did not have much variability in age as they were all undergraduate students, and we did not expect that the age differences would be a differential contributor to the study findings. In response to the reviewers, we calculated the correlation between age and state and trait anxiety and adherence, and none of the results were significant (trait anxiety: r=0.109; P=.48; state anxiety: r=0.125; P=.37; adherence: r=−0.015; P=.91).

It may be useful to make explicit the assumptions underlying the modeling and parameters used for PROCESS, such as the degree of independence of the moderator.

Response: The parameters for using moderation in PROCESS are like any moderation analysis program. The only difference is the bootstrapping, which is specified in the Methods section. We now added this information in the manuscript and the new reference as well.

Hayes AF. *Introduction to Mediation, Moderation, and Conditional Process Analysis.* 3rd Edition. Guilford Press, 2022.

Discussion

The authors may consider adding a section to the discussion to explore variables related to vaccine hesitancy and other factors (eg, sense of invincibility) as a suggestion for future research, expanding the scope beyond adherence to preventive measures.

Response: We agree with this comment, and we expanded on this topic with additional sentences added in the Discussion section.

Given the reliance on self-report measures, the reviewers suggest the authors discuss the potential impact of social desirability bias on participants' responses. Addressing this concern would add transparency to the limitations of the study.

Response: We agree with this comment. We wanted to point out that we had already addressed this issue in the Limitation section.

Reviewers suggest authors discuss how the results support following up further on correlations among PTM scales and on the possible moderator effect of public prosocial tendencies, with recommendations for including a broader set (explicitly listed) of potentially explanatory independent variables.

Response: We agree with the reviewer's comments. We now added a list of potential variables that may have also influenced the results in the Discussion section. We added the following paragraph: "Additionally, other variables that may have influenced the results such as vaccine hesitancy, perceived personal risk and disease vulnerability, and trust in science may be potential variables to study in future research, especially regarding the factors that may impact the adherence to preventive behaviors in young adults [7,8]."

It may also be helpful to add some explanation of why the psychometric characteristics of the survey instruments as established in other studies can be trusted to be the same as used in this study (online, unsupervised, etc). Some reviewers found it concerning that this study found statistically significant pairwise associations between PTM subscales, and this should be addressed, perhaps with speculation about why this happened.

Response: We thank the reviewers for this insightful comment. We now added the validity and reliability information of all the PTM subscales and of the State-Trait Anxiety Inventory. We also further elaborated on the possible reasoning for the differential results of the PTM in the Discussion section.

Figures and Tables

Consider using a 2 × 2 table in Figure 1 to illustrate the detected moderator effect.

Response: We thank the reviewers for this comment. However, because of the many tables we already have in the study and the fact that we added an additional one with the demographics (Table 1), we decided to use this figure that illustrates the study's effects without having to have 2 figures.

Title

Given the concern about generalizability, a reviewer suggested the authors consider changing the title to "Adherence Behaviors to Prevent COVID: The Role of Anxiety and Prosocial Behaviors Amongst University Students in the US."

Response: We thank the reviewers for this comment, and we also considered that it would be important to add the dates of the study to describe the historical circumstances; therefore, we also added dates to increase awareness of the time period. Therefore, the manuscript is now titled "Adherence Behaviors to Prevent COVID: The Role of Anxiety and Prosocial Behaviors Amongst University Students in the US- January 2021-May 2021."

References

1. Arogundade FQ, Azra S, Pulier M, Ram L, Saderi D, Tomaskova M. Peer review of "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study". JMIRx Med 2024;5:e59430. [doi: 10.2196/59430]
2. Corbera S, Marin-Chollom AM. The role of anxiety and prosocial behaviors on adherence behaviors to prevent COVID-19 in university students in the United States: cross-sectional study. JMIRx Med 2024;5:e52970. [doi: 10.2196/52970]
3. Starcke K, Brand M. Decision making under stress: a selective review. Neurosci Biobehav Rev 2012 Apr;36(4):1228-1248. [doi: 10.1016/j.neubiorev.2012.02.003] [Medline: 22342781]

4. Nowacki J, Heekeren HR, Deuter CE, et al. Decision making in response to physiological and combined physiological and psychosocial stress. Behav Neurosci 2019 Feb;133(1):59-67. [doi: 10.1037/bne0000288] [Medline: 30556701]
5. Rudenstine S, McNeal K, Schulder T, et al. Depression and anxiety during the COVID-19 pandemic in an urban, low-income public university sample. J Trauma Stress 2021 Feb;34(1):12-22. [doi: 10.1002/jts.22600] [Medline: 33045107]
6. Lee J, Solomon M, Stead T, Kwon B, Ganti L. Impact of COVID-19 on the mental health of US college students. BMC Psychol 2021 Jun 8;9(1):95. [doi: 10.1186/s40359-021-00598-3] [Medline: 34103081]
7. Gupta S, Watanabe S, Laurent SM. Psychological predictors of vaccination intentions among U.S. undergraduates and online panel workers during the 2020 COVID-19 pandemic. PLoS One 2021 Nov 30;16(11):e0260380. [doi: 10.1371/journal.pone.0260380] [Medline: 34847162]
8. Hromatko I, Tonković M, Vranic A. Trust in science, perceived vulnerability to disease, and adherence to pharmacological and non-pharmacological COVID-19 recommendations. Front Psychol 2021 Apr 30;12:664554. [doi: 10.3389/fpsyg.2021.664554] [Medline: 33995222]

Abbreviations

IRB: institutional review board
PTM: Prosocial Tendencies Measure

Edited by T Leung; submitted 26.03.24; this is a non–peer-reviewed article;accepted 26.03.24; published 27.05.24.

Please cite as:
Corbera S, Marín-Chollom AM
Authors' Response to Peer Reviews of "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study"
JMIRx Med 2024;5:e58859
URL: https://xmed.jmir.org/2024/1/e58859
doi:10.2196/58859

Authors' Response to Peer Reviews of "Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis"

Tim Dong[1], MSc; Shubhra Sinha[1], MBBS; Ben Zhai[2], PhD; Daniel Fudulu[1], MD, PhD; Jeremy Chan[1], MD; Pradeep Narayan[3], MD; Andy Judge[1], PhD; Massimo Caputo[1], MD; Arnaldo Dimagli[1], MD; Umberto Benedetto[1], MD, PhD; Gianni D Angelini[1], MD

1

2

3

Corresponding Author:
Tim Dong, MSc

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45973

Companion article: https://www.medrxiv.org/content/10.1101/2023.01.21.23284795v1

Companion article: https://med.jmirx.org/2024/1/e60428

Companion article: https://med.jmirx.org/2024/1/e60280

Companion article: https://med.jmirx.org/2024/1/e45973

(JMIRx Med 2024;5:e60384) doi:10.2196/60384

KEYWORDS

cardiac surgery; artificial intelligence; risk prediction; machine learning; operative mortality; data set drift; performance drift; national data set; adult; data; cardiac; surgery; cardiology; heart; risk; prediction; United Kingdom; mortality; performance; model

This is the authors' response to peer-review reports for "Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis."

Round 1 Review

Anonymous [1]

General Comments

Overall, I think this is a really interesting paper [2]. It is a concept I had never heard of, and I can see very clearly how this is an important consideration. I also think the authors have done excellently to consider a host of different aspects, including feature importance change, beyond the most obvious measurements.

Specific Comments

Abstract

1. "It has been suggested that using Machine Learning (ML) techniques, a branch of Artificial intelligence (AI), may improve the accuracy of risk prediction." Improve them over what? Specify what the status quo is with regard to first principles and data-driven modeling. This statement is also repeated in the first line of the introduction—what is "conventional" about these models?

Response: Please note that all line numbers refer to the marked version of the manuscript with tracked changes, uploaded as a supplementary file.

Thank you for this helpful suggestion. This has now been modified in the abstract as follows:

> *It has been suggested that using Machine Learning (ML) techniques, a branch of Artificial intelligence (AI), may improve the accuracy of risk prediction over traditional mortality risk stratification models.*
>
> *These traditional scoring methods are generally based on logistic regression with risk factors determined through a consensus across experts within leading cardiac surgery organisations in the United States (STS) or Europe (ES II).*

The above important points have been incorporated in lines 31, 103, and 106-108.

2. "five ML mortality prediction models"—it should be highlighted that these are novel models that you have developed for this paper.

Response: Thank you for this suggestion. This has now been modified as "Five novel ML mortality prediction models were developed and assessed with EuroSCORE II for relationships…" in lines 39-40.

3. "geometric average results of all metrics"—it is not all metrics, just the 5 that you have calculated. It is better to just say here "a novel metric called the CEM" or something.

Response: Thank you for the helpful suggestion. This has been changed to "Performance was assessed using a consensus metric." in lines 41-44.

Introduction

Why is data set drift a problem? I think you could do more here to highlight how important this is to an audience who might not be dealing with the data themselves and, thus, might not naturally think of examples: for example, changes in treatment guidelines, demographics, new risk factors emerging, or changes in coding practices. You could mention "new" comorbidities such as long COVID.

Response: Thank you for this interesting comment. We have now included a more extensive explanation of data set drift and its importance in the *Introduction* section.

> *Introduction*
>
> *Changes in treatment regimens, demography, new risk factors, adjustments to clinical coding procedures, or the addition of new variables such as the identification of previously unknown conditions such long Covid can all contribute to this phenomenon. The issue of dataset drift is serious, particularly for individuals who depend on the quality or insights of the data but may not analyse it directly. Below are some reasons why this is important:*
>
> *1. Impact on Decision-Making: Decision-makers may base their choices on erroneous or obsolete information if they rely on drifted or outdated datasets. In the healthcare industry, for example, if changes in treatment recommendations rely solely on historical data, this could result in less-than-ideal patient care because the analysis would not account for newer, more effective therapies.*
>
> *2. Reduced Model Performance: When dataset drift occurs, machine learning models and predictive algorithms that were trained on historical data may become less accurate or dependable. For instance, a financial prediction model based on antiquated market tendencies may not be able to predict novel market behaviours, which could result in losses.*
>
> *3. Biased or Inaccurate Insights: Datasets that have drifted may contain biases or errors. Model generalizability may be impacted by changes in demographics, such as adjustments in the age distributions of the population. The prevalence of post-cardiac surgery outcomes may be impacted by newly identified risk factors or circumstances (such as long Covid), necessitating modifications to predictive models in order to preserve accuracy.*
>
> *4. Challenges in Generalization: Models developed with dated data could have trouble extrapolating to novel scenarios. For instance, Euroscore I was developed in 1999 using 19,030 patients collected over three months (September–December 1995) from 132 cardiac centres in eight countries [13]. Modifications to risk factors over time may have an resulted in its lack of discrimination and calibration compared to its successor score EuroSCORE (ES) II, developed in 2011.*
>
> *5. Ethical and Fairness Concerns: Drift in the dataset may exacerbate problems with ethics or fairness. A system may reinforce preexisting biases and unfairly target particular groups if it was trained on biased or out-of-date data.*
>
> *6. Regulatory Compliance: Using historical or drifted data could result in non-compliance with changing standards that need accurate, current information in regulated industries like healthcare.*

The above important points have been incorporated in lines 148-181.

Methods

1. Could the same individuals be in both the training and validation set and holdout set, if they had multiple surgeries? If so, this may have introduced some bias into the performance estimates. I do not think you need to redo the analyses, but if you can highlight the degree of overlap, then that would be good. Otherwise, say it was not possible and list it as a limitation.

Response: Thank you. As National Adult Cardiac Surgery Audit (NACSA) patient identifiers and the Hospital Episode Statistics data set were not available for linkage, it was not possible to determine whether there were any patients in both the training and validation set and holdout set, where they had multiple surgeries. Clinical judgment suggests that the proportion with multiple surgeries would be very low. Nonetheless, future work should consider the collection of such information to minimize any potential bias.

The above important points have been incorporated in lines 644-649.

2. "As a sensitivity analysis, we excluded the True Negative Rate from the performance evaluation, by calculating the F1 score." This sentence does not quite make sense to me. The F_1-score is based on the sensitivity (true negative rate) and the precision (positive predictive value), right? It does not exclude the true negative rate per se; it just does not use it.

Response: Thank you. As a sensitivity analysis, we calculated the F_1-score, which combines precision and recall without explicitly considering the true negative rate in the performance evaluation.

The above important points have been incorporated in lines 274-277.

Thank you for your expert and invaluable review; this has been really thought provoking and had made significant contributions to improving the journal's quality of output. Thanks.

We hope this now addresses your queries. Thank you.

Reviewer CL [3]

General Comments

This manuscript presents an interesting study that explores temporal trends in various performance metrics for different types of prediction models used in the prediction of in-hospital mortality after cardiac surgery in the United Kingdom from 2012 to 2019. The data set was divided into 2 periods: from 2012 to 2016 for model training and internal validation and from 2017 to 2019 for external validation. The study evaluated 5 prediction models: logistic regression, support vector machine (SVM), random forest, extreme gradient boosting (XGBoost), neural network, and European System for Cardiac Operative Risk Evaluation (EuroSCORE) II. The authors aimed to assess the model performance on 5 metrics (1 – expected calibration error [ECE], area under the curve [AUC], 1 – Brier score, F_1-score, and net benefit) and proposed a composite metric, the clinical effectiveness metric (CEM), calculated as the geometric mean of the 5 mentioned metrics, as the primary metric.

The study began with a nontemporal baseline evaluation of different models in the 2017-2019 temporal validation and then conducted a series of drift analyses, including an examination of overall trends from 2012 to 2019, within-period trends in the first 3 months of 2017 and 2019, and between-period trends between the first 3 months of 2017 and 2019. The authors also analyzed drift in variable importance and variable distribution, defined by the temporal change in the ratio of several top-importance features within the data set, to profile data set drift.

The authors demonstrated that XGBoost and random forest were the best-performing models, both in nontemporal and temporal evaluations, whereas the EuroSCORE II model exhibited a significant drop in performance. Temporal declines in model performance were observed across all models and were consistent with data set drift.

Overall, the question of the generalizability of prediction models, whether temporal or spatial, has long been a topic of discussion in clinical research. This study takes a commendable approach to addressing this question. However, there are some issues that require clarification and revision, including (1) methodological concerns related to the justification of the main metric (CEM) using averaging, and the appropriateness of some statistical tests; (2) the clinical significance of the identified performance drift; and (3) the overall clarity of the study's design and presentation.

Specific Comments

Major Comments

1. The statement of the study's objectives should be improved for more clarity, particularly regarding the phrase "verify suspected dataset drift by assessing the relationship between and within performance drift, variable importance drift, and dataset drift across ML and ES II approaches." It is unclear what is meant by the "relationship between and within." Does this refer to the analysis of performance drift within and between different periods? The overall study design is quite challenging to grasp initially, even with the graphical overview provided in Figure 1. To enhance clarity, additional details and explanations should be added to the aims, overall design, graphical overview, and text the Methods and Results sections.

Response: Please note that all line numbers refer to the marked version of the manuscript with tracked changes, uploaded as a supplementary file.

Thank you for this helpful suggestion. We have now improved the *Introduction* and *Methods* sections in terms of making the design and aims easier to comprehend across a wider range of readers.

> *Introduction*
>
> *Performance drift in ML is when the performance of the ML models deteriorate over time due to various changes that may reduce the validity of the model's assumptions at the time of training. The following are the primary reasons for performance drift: (i) Dataset Drift happens when the distribution of the data between the training set and the dataset used for evaluation or prediction varies. For example, if a model is trained on data from one time period but evaluated or used in another time period where the data distribution has changed dramatically, performance may suffer; (ii) Concept Drift occurs when the fundamental relationship between the input features and the target variable shifts over time. The assumptions upon which the model was developed may no longer be valid. In a predictive maintenance model, for example, the behavior of the machinery may alter subtly over time due to numerous causes (such as wear and tear), causing the model to become less accurate as time passes; (iii) Variable Importance Drift: Changes in the significance or importance of various variables/features used by the model to make predictions. Variables that were essential during the model's training phase may become less important, while other variables may become more influential when the environment or the problem itself evolves. Calibration Drift: Calibration refers to the agreement between expected and actual probabilities of an event; (iv) Calibration drift occurs when the estimated probability of the model grow less dependable over time. This could happen if the model was calibrated based on assumptions about the data distribution that no longer hold true. These different types of drift may also have an interplay effect, and this was shown through a non-cardiac surgery study that used actual dataset drift to verify variable importance detected dataset drift [54].*
>
> *Changes in treatment regimens, demography, new risk factors, adjustments to clinical coding procedures, or the addition of new variables such as*

the identification of previously unknown conditions such long Covid can all contribute to this dataset drift phenomenon. The issue of dataset drift is serious, particularly for individuals who depend on the quality or insights of the data but may not analyse it directly. Below are some reasons why this is important:

1. Impact on Decision-Making: Decision-makers may base their choices on erroneous or obsolete information if they rely on drifted or outdated datasets. In the healthcare industry, for example, if changes in treatment recommendations rely solely on historical data, this could result in less-than-ideal patient care because the analysis would not account for newer, more effective therapies.

2. Reduced Model Performance: When dataset drift occurs, machine learning models and predictive algorithms that were trained on historical data may become less accurate or dependable. For instance, a financial prediction model based on antiquated market tendencies may not be able to predict novel market behaviours, which could result in losses.

3. Biased or Inaccurate Insights: Datasets that have drifted may contain biases or errors. Model generalizability may be impacted by changes in demographics, such as adjustments in the age distributions of the population. The prevalence of post-cardiac surgery outcomes may be impacted by newly identified risk factors or circumstances (such as long Covid), necessitating modifications to predictive models in order to preserve accuracy.

4. Challenges in Generalization: Models developed with dated data could have trouble extrapolating to novel scenarios. For instance, Euroscore I was developed in 1999 using 19,030 patients collected over three months (September–December 1995) from 132 cardiac centres in eight countries [13]. Modifications to risk factors over time may have an resulted in its lack of discrimination and calibration compared to its successor score EuroSCORE (ES) II, developed in 2011.

5. Ethical and Fairness Concerns: Drift in the dataset may exacerbate problems with ethics or fairness. A system may reinforce preexisting biases and unfairly target particular groups if it was trained on biased or out-of-date data.

6. Regulatory Compliance: Using historical or drifted data could result in non-compliance with changing standards that need accurate, current information in regulated industries like healthcare.

The aim of this study was to investigate performance drift in existing ML models that have been used in prior cardiac surgery risk prediction research. The objectives were to (i) rank and assess the extent of performance drift in such cardiac surgery risk ML models over time; (ii) investigate any potential influence of dataset drift and variable importance drift on performance drift.

We have also added an extensive *Related Work* section in the *Introduction* section.

Future Work

Future studies shall also delve deeper into the relationships of the studied drift types with concept drift in cardiac surgery risk prediction.

The above important points have been incorporated in lines 123-186, 194-217, and 667-668.

2. The rationale for introducing CEM as the primary performance metric, calculated as the geometric mean of 5 distinct individual metrics, is debatable and lacks strong justification. Although the geometric mean is less sensitive to outliers compared to the arithmetic mean, it raises the fundamental question of why these metrics need to be summarized. Is it merely to obtain a single quantitative measure for analysis, or does it aim to provide a more comprehensive understanding of overall model performance? It appears to serve primarily the former purpose, which may not be an appropriate practice given that the 5 metrics assess entirely different aspects of model performance: 1 – ECE for calibration, AUC for discrimination, 1 – Brier score (which already encompasses calibration and discrimination components), F_1-score for threshold-specific discrimination, and net benefit index for cost-effectiveness. Consequently, interpreting the exact meaning of CEM becomes challenging, as it reduces these diverse aspects to a single numerical value. Therefore, I suggest just reporting and examining all 5 metrics individually, with or without highlighting certain ones as primary areas of interest.

Response: Thank you for your comment. In our previous study, we found that combining the metrics covering all 4 aspects of discrimination, calibration, clinical usefulness, and overall accuracy into a single CEM improved the efficiency of cognitive decision-making (according to the Miller Law [4]) for selecting the optimal ensemble models [5,6]. This approach is useful for providing a consensus metric that enables models to be ranked in scenarios where, for example, 1 model could outperform another using 1 metric but underperform under a different metric. Furthermore, we demonstrated that such a consensus metric could be combined with drill-down analysis to further interpret the models using individual metrics [5]. Although AUC evaluates the diagnostic or predictive performance of the model, it does not directly reflect patient benefit. This is why we included a suit of other metrics, including the decision curve analysis net benefit index, that were found to be clinically pertinent from our prior study [7].

The above important points have been incorporated in lines 195-205.

3. The manuscript used several statistical tests, and some of them are relatively less commonly used. Please provide a more detailed description of the objectives and specific statistical situations for each test used. Additionally, for the baseline nontemporal performance comparison, a more conventional approach for comparing AUC would be the use of the DeLong method (you could choose the best model as the reference), and bootstrapping can be used to assess the statistical significance when comparing other metrics.

Response: Thank you. A list of statistical methods used for analyzing drift has been provided in Table 1 (please see the manuscript for the formatted version, thanks).

> *Objective Statistical Tests General Statistical Situations Rationale for Choosing Test Assumptions Checked*
>
> *Non-temporal comparison of models Repeated measures One-Way ANOVA Comparison of multiple groups for differences Used for comparing means across multiple models Outliers (ANOVA assumptions), Normality (Shapiro-Wilk test)*
>
> *Paired t-tests (Bonferroni Corrected) Comparison of paired observations between models To compare specific model pairs*
>
> *Dunnett's Correction Control for multiple comparisons Controls Type I error rate in comparing multiple treatments to a control group in one-way ANOVA*
>
> *Analysis within specific time frames Kruskal-Wallis Test Comparison of multiple groups for differences (non-parametric) Non-parametric alternative for ANOVA in specific time frames Outliers (ANOVA assumptions), Normality (Shapiro-Wilk test)*
>
> *Bonferroni Corrected Paired samples Wilcoxon test (Wilcoxon signed-rank test) Comparison of paired observations within time frames Non-parametric comparison of paired samples within time frames with control for Type I error rate in comparing multiple treatments*
>
> *Dunn's test Multiple pairwise comparisons within non-parametric groups Post hoc test for pairwise comparisons after Kruskal-Wallis test; Determines the magnitude of difference effects within time frames*
>
> *Analysis between first 3 months of 2017 and 2019 Kruskal-Wallis Test Comparison of multiple groups for differences (non-parametric) Non-parametric comparison between time frames Outliers (ANOVA assumptions), Normality (Kolmogorov-Smirnov Test)*
>
> *Paired samples Wilcoxon test (Wilcoxon signed-rank test) Comparison of paired observations between time frames Non-parametric comparison of paired samples between time frames*
>
> *Bonferroni adjusted Dunn's test Multiple pairwise comparisons between time frames Post hoc test for pairwise comparisons after significant Kruskal-Wallis results; Determines the magnitude of difference effects between time frames; with control for Type I error rate in comparing multiple treatments*
>
> *Normality (Kolmogorov-Smirnov Test)*
>
> *Analysis of discrimination, calibration, clinical utility, and overall accuracy drift Linear regression (with residual analysis) Assessing relationships and regression parameters To analyze linear relationships and model residuals Normality through histograms and QQ plots,*
>
> *Seasonal Kendall Test (Non-parametric alternative if assumptions not met) Assessing association or trends when assumptions are not met Non-parametric test for assessing associations without assumptions Homoscedasticity through scale-location plots*
>
> *We appreciate the suggestion regarding DeLong's method for assessing AUC comparison. Future study could investigate the utility of DeLong's method in measuring AUC differences, particularly in studies focusing on pairwise model comparisons. The computational demands of this strategy, which can be burdensome on large datasets, impacted our decision not to use it in the current study. However, given its proven importance in AUC comparisons, future studies with a focus on AUC evaluation and resource availability for controlling computational demands may explore using DeLong's method. This method could aid in the refinement of comparison analyses in predictive modelling research by allowing for a more complete knowledge of AUC differences between models.*
>
> *We wanted to analyse model performance across multiple metrics across time in this study. Although DeLong's approach is often used for pairwise comparisons of the area under the curve (AUC), we chose not to utilise it due to its high computational demands [58], particularly on the large datasets present in this study. We chose a more comprehensive approach to capture the dynamics of model performance since this study included a broad examination across various performance metrics rather than focusing solely on AUC.*

The above important points have been incorporated in lines 308-309, 580-586, and 659-665.

4. During the training and internal validation phase with 5-fold cross-validation, additional details are needed to understand how the final model for each model type was selected for subsequent temporal validation, including whether hyperparameter tuning was carried out and whether there was a final refitting process on the entire training data set following the cross-validation, etc.

Response: Thank you. Internal validation was performed using 5-fold cross-validation on the training and validation data set (2012-2016) to select model parameters. The final models were determined by retraining the models on the combined training and validation data set using the selected model parameters. Temporal validation was performed using the final models on the holdout data set (2017-2019) [8]. Further details on model development can be found in the *Model Specification* section in Multimedia Appendix 1.

> *Supplementary Section 2: Model Specification*
>
> *Neural Network (Neuronetwork) was trained using 1000 epochs, with batch size of 20,000. The 2012-2016 dataset was split 70:30, with 70% used as training data and 30% as validation data for early stopping to reduce likelihood of overfitting [1]. The best model was saved using early stopping to prevent*

overfitting [2]. Binary cross-entropy loss was used as the loss function, with Adam as the optimizer [3], monitoring on accuracy as the metric. The final model configuration used for evaluation was the optimal set derived from the NACSA Bristol cohort from our previous study: input layer n=18 nodes, hidden layer one n=90 nodes, hidden layer two n=36 nodes and output layer one node [4].

3-fold Grid Search Cross Validation was applied for Weighted SVM and Xgboost using 2012-2016 dataset to determine the optimal hyperparameters to apply to 2017-2019 test dataset [5]. For Random Forest, we manually tuned parameters in response to model discrimination (AUC) evaluated with cross-validation (estimators n=700, maximum depth n=10, minimum samples split n=5, minimum samples leaf n=20) [4]. The ES II risk factors were fitted with an LR (retrained LR) model with Inverse of regularization strength (C) set to 1 [4].

The above important points have been incorporated in lines 258-261.

5. The Introduction section should incorporate more background information on previous studies reporting or relating to performance variation in prediction models for cardiac surgery outcomes. In the Discussion section, it is also important to discuss how this work contributes to existing evidence in the context of these previous studies. Some relevant studies, based on my preliminary search, include Benedetto et al [9], Zeng et al [10], Mori et al [11], and potentially more.

Response: Thank you. We have now included a *Related Works* section to enhance the *Introduction* section.

Related Work

In our previous study, we found that combining the metrics covering all four aspects of discrimination, calibration, clinical usefulness and overall accuracy into a single CEM improved the efficiency of cognitive decision-making (according to Miller's Law [16] for selecting the optimal ensemble models [13,17]. This approach is useful for providing a consensus metric that enables models to be ranked in scenarios where for example one model could outperform another using one metric, but underperform under a different metric. Furthermore, we demonstrated that such a consensus metric could be combined with drill-down analysis to further interpret the models using individual metrics [13]. While AUC does evaluate diagnostic or predictive performance of the model, it does not directly reflect patient benefit. This is why we had included a suit of other metrics including the Decision Curve net benefit index that were found to be clinically pertinent from our prior study [18].

In our previous work [19], we had studied the calibration changes across two different time intervals using the calibration belt (overall external calibration) and Hosmer–Lemeshow goodness of fit χ^2 statistics (calibration drift) approach within a single United Kingdom based hospital. A recent study extended our work to a Chinese national registry, Sino (Chinese) System for Coronary artery bypass grafting (CABG) Operative Risk Evaluation II (SinoSCORE II), using an set of ML models included lightGBM, CatBoost and a combination of variable selection approaches including Optuna for stepwise regression (SWR), BorutaSHAP (BS), and feature importance (FI) ranking [20]. Another study in the United States (U.S) had also investigated the calibration performance difference between Xgboost and Logistic Regression models built for the CABG patient cohort through pre-operative, intra-operative and combined variable sets from the Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database (ACSD) [21].

We have discussed these in the *Discussion* section where appropriate.

Discussion

Our previous study [19], while not involving the assessment of xgboost had also shown that calibration drift of Logistic Regression was less than that of Random Forest, while EuroSCORE I, Naïve Bayes and Neural Network performed poorly in terms of calibration. A recent study extending upon our work had shown that temporal and spatial calibration drift (comparison across regions and hosptials) to be severe across a range of ML models using a national Chinese registry [20]. In accordance with our view, the study highlighted that "future efforts may need to shift more towards enhancing model calibration robustness or recalibration for greater practical value" and that inclusion of intra-operative variables may be important to enhancing model performance. This Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database (ACSD) study [21], had shown that the inclusion of intra-operative variables improved both the discrimination and calibration performance of Xgboost and Logistic Regression models in CABG patients from the U.S.

The above important points have been incorporated in lines 195-217 and 560-572.

6. Although the authors observed numerical declines in CEM and other metrics, the magnitude of these declines appears to be relatively small, particularly when considering metrics such as AUC. As a result, it is essential to discuss how to interpret this magnitude of drift in the context of clinical practice. In other words, what is the clinical significance of this variation in performance, and how does it justify the necessity of actively monitoring model drift in terms of cost-effectiveness? Please discuss.

Response: Thank you for this interesting suggestion. We have now expanded the *Discussion* section to further discuss this suggested topic.

Discussion

Although the reported decreases in measures such as CEM and AUC may appear small, such changes are likely to impact the potential usage of ML models within clinical scenarios. If such models are to be used clinically for making decisions about the patient, even small changes in these metrics (which have been previously discussed[18] to be important in the cardiac surgery ML performance) can have an influence on risk assessment and patient outcomes, necessitating constant model drift monitoring. Prior research has shown that improving model calibration robustness or recalibration is necessary for practical value and that the "the significant decline in performance of previously established models in this study calls for continuing model updates"[20]. It is envisaged that collaboration between physicians and ML scientists is critical. Before mandating model updates, it is critical to establish metric-specific thresholds for acceptable reductions. A consensus approach, extensive experience in this area or a meta-analysis of current literature may be required for this collaborative decision-making process.

The above important points have been incorporated in lines 612-624.

7. The conclusion should only focus on the primary findings outlined in the aims of the Introduction section. Avoid incorporating less central findings and speculative elements. Additionally, it may not be fair to suggest replacing the EuroSCORE II model simply based on the inferior performance in this study, since it was already established and this study essentially conducted an external validation for it, whereas the other machine learning models were developed using these data sets.

Response: Thank you for this valuable suggestion. We have now revised the *Conclusion* section to make this more coherent and focused.

Conclusion

This study found that performance drift of ML and ES II over time could be explained through dataset drift patterns in cardiac surgery risk prediction. It was also found that variable importance drift could help to explain performance drift and support detection of dataset drift in the assessed models. The strong evidence of all models showing a decrease in at least 3 of the 5 individual metrics within CEM demonstrates the potential need to update the models over time but future work are required to determine suitable thresholds for mandating an update. Future work will be required to determine the interplay between Xgboost and RF, which have demonstrated less drift over time, and whether combining these through additional ensemble modelling could take advantage of their respective performance advantages.

The above important points have been incorporated in lines 671-690 and 534-535.

Minor Comments

1. More detailed definitions and explanations should be provided for each performance metric.

Response: Thank you for this helpful comment. This has now been included as part of the *Related Work* section.

Related Work

In our previous study, we found that combining the metrics covering all four aspects of discrimination, calibration, clinical usefulness and overall accuracy into a single CEM improved the efficiency of cognitive decision-making (according to Miller's Law[16]) for selecting the optimal ensemble models [13,17]. This approach is useful for providing a consensus metric that enables models to be ranked in scenarios where for example one model could outperform another using one metric, but underperform under a different metric. Furthermore, we demonstrated that such a consensus metric could be combined with drill-down analysis to further interpret the models using individual metrics [13]. While AUC does evaluate diagnostic or predictive performance of the model, it does not directly reflect patient benefit. This is why we had included a suit of other metrics including the Decision Curve net benefit index that were found to be clinically pertinent from our prior study [18].

The above important points have been incorporated in lines 195-205.

2. In the Methods section, please provide a clear outline of the inclusion and exclusion criteria. Additionally, consider including a flowchart that illustrates the data set development process, outlining how these criteria were applied.

Response: Thank you. We have now updated the *Methods* section and Multimedia Appendix 1.

Methods

227,087 adults patients undergoing cardiac surgery between January 1, 2012 and March 31, 2019 were included. Congenital, transplant and mechanical support device insertion cases were excluded. A patient flow consort diagram is shown in Supplemental materials, Figure S1.

Supplementary Materials

Consort Diagram

Figure S1.1 Consort diagram showing flow of participants through the study.

The above important points have been incorporated in lines 224-227 and Figure S1 in Multimedia Appendix 1.

3. I had difficulty understanding what "outliers" and "distribution" meant in the Results section for the baseline nontemporal performance of each model. I thought that each metric of each model should be just a numerical value and a 95% CI from bootstrapping.

Response: Thank you for this helpful suggestion. We have now improved clarity of the *Results* section (see underlined parts below).

> *Baseline non-temporal performance*
>
> *No extreme outliers were found when testing for ANOVA assumptions. The CEM scores from 1000 bootstraps were normally distributed for all three models except Xgboost,*

The above important points have been incorporated in lines 368-369.

4. The title of the manuscript should be an objective reflection of the overall study design and aim, rather than drawing conclusions from the findings.

Response: Thank you for the valuable suggestion. We have now amended the title to "An assessment of performance drift in Machine Learning models for cardiac surgery risk prediction."

The above important points have been incorporated in lines 2-4.

5. I did not find the supplementary materials in the review system. I am not sure whether this issue is on my end or not.

Response: Thank you. We have now reuploaded the latest changes in the supplementary materials to the journal upload page. We have also included updated figures within the supplementary materials with our responses wherever possible.

Thank you for your expert and invaluable review; this has been really thought provoking and had made significant contributions to improving the journal's quality of output. We hope our changes will be met with your approval.

Round 2 Review

Reviewer CL

General Comments

I appreciate the opportunity to rereview this manuscript. The authors' efforts in revising their manuscript in response to previous concerns are commendable. This manuscript has been improved and is now in principle publishable. It could potentially be accepted upon reasonable response to a few follow-up minor comments, outlined below.

Specific Comments

1. About my previous major comment 1, the authors meticulously elaborated on (1) the reasons for performance drift and (2) its importance, which are both valid points. However, the current Introduction (lines 121-179) is quite lengthy. I recommend consolidating these 2 parts into a single paragraph, listing each point without the need for detailed individual explanations. Additionally, my query about the exact meaning of "the relationship between and within variable importance drift, performance drift, and actual dataset drift" remains unaddressed. Even though it was removed from the Introduction, it still appears in the abstract. I suggest the authors explicitly explain it to readers and incorporate it into the manuscript when first mentioned.

Response: Please note that all line numbers refer to the marked version of the manuscript with tracked changes, uploaded as a supplementary file.

Thank you for this helpful suggestion. We have now improved the *Introduction* section by making it more concise and explaining the meaning of "the relationship between and within variable importance drift, performance drift, and actual dataset drift" as per your recommendations:

> *Introduction*
>
> *In machine learning (ML), performance drift refers to the gradual loss in model performance caused by changes that call into question the model's training assumptions. Key causes of performance drift include dataset drift, which refers to changes in the distribution of data between training and evaluation sets; variable importance drift, which involves changes in the significance of model variables; and calibration drift, which is characterised by decreased reliability in estimated probabilities. These factors can interact, as seen in a study of non-cardiac surgery [54]. Understanding the complex relationship between variable importance drift, performance drift, and dataset drift is important. This relationship explains how changes in the importance of specific variables, combined with changes in the actual data distribution, collectively influence the model's overall accuracy and reliability as it performs over time. The wider implications are also significant, influencing decision-making, insight accuracy, generalisation [13], ethical considerations, and regulatory compliance across industries.*

The above important points have been incorporated in lines 121-192.

2. Regarding the justification for the CEM, the authors have added more explanation and supporting literature for its use. However, it would strengthen their case if they could provide examples from external studies or use cases where a similar practice (averaging different aspects of metrics for model performance evaluation) was used, beyond their own studies.

Response: Thank you for your comment. We have now strengthened the case by providing examples of additional external studies where a similar practice have been applied:

> *Methods*
>
> *The consensus approach for combining different metrics has previously been applied in a study on Covid-19 prediction [33]. In addition, this approach is similar to the simple additive weighting (SAW) multi-criteria evaluation approach for making a decision through the ranking of a set of competing criterions [34].*

The above important points have been incorporated in lines 294-298.

3. About the statistical tests for comparing AUC with the DeLong method, I believe that performing the DeLong test for AUC comparison is not overly computationally demanding, even on a relatively large data set. I recommend the authors explore commonly used R packages (eg, "pROC") that facilitate AUC calculation and comparison with the DeLong method. The DeLong comparison typically requires paired variables of the

label and 2 models' predicted probabilities, and the 95% CI and P value are automatically calculated by bootstrapping these paired samples, which is relatively efficient.

Response: Thank you for further explaining. We have know included the DeLong test in the baseline nontemporal comparison as you advised.

> *Baseline non-temporal performance (methods)*
>
> *The Delong's test was applied for determining whether there was a statistically significant difference across the AUCs of ROC curves for the top two best performing models.*
>
> *Baseline non-temporal performance (results)*
>
> *AUC performance was best for Xgboost (0.834) and RF (0.835), with the Delong's test showing no statistically significant difference (P>.05).*
>
> *Table below has been updated to include the Delong's test:*
>
> *Table 1a. Summary of statistical methods used for assessing drift.*
>
> *Objective Statistical Tests General Statistical Situations Rationale for Choosing Test Assumptions Checked*
>
> *Non-temporal comparison of models Repeated measures One-Way ANOVA Comparison of multiple groups for differences Used for comparing means across multiple models Outliers (ANOVA assumptions), Normality (Shapiro-Wilk test)*
>
> *Paired t-tests (Bonferroni Corrected) Comparison of paired observations between models To compare specific model pairs simultaneously*
>
> *Dunnett's Correction Control for multiple comparisons Controls Type I error rate in comparing multiple treatments to a control group in one-way ANOVA*
>
> *Delong's test Comparison of the AUC of two correlated ROC curves To compare AUC of two models/tests during sensitivity testing*

The above important points have been incorporated in lines 316-318 and 398-399 and Table 1.

4. Regarding model tuning and specification of the best models (PS: I still cannot find the supplements, only a revised clean manuscript; I am not sure if this was due to issues from my end), I am curious why different tuning practices were used for different models, especially grid search for XGBoost and SVM but manual tuning for random forest.

Response: Thank you for your helpful comment. For random forest, the final model configuration used for evaluation was the optimal set derived from our previous study on the NACSA Bristol cohort [9]. For the new models SVM and XGBoost, for which optimal parameters have not been investigated in our previous study [9], we applied 3-fold grid search cross-validation to determine the optimal hyperparameters.

The above important points have been incorporated by updating the *Model Specification* section in Multimedia Appendix 1.

5. In response to the query about the clinical significance of the relatively small scale of performance drift, the authors referred to one of their previous studies briefly discussing this matter. However, it would be much clearer if the authors could more explicitly elaborate in this study and, if possible, provide additional analysis to support this argument.

Response: Thank you for this invaluable comment.

> *Net benefit projection (methods)*
>
> *To further understand the clinical significance of the performance drift over time, the fitted linear regression model intercepts and slopes was used to extrapolate the net benefit up to January 2030 for Xgboost and Neural Network models.*
>
> *Net benefit projection (results)*
>
> *To further understand the clinical significance of the performance drift over time, Figure 5 illustrates the expected net benefit decrease for a NN model and an XGBoost model. The blue line depicts the actual net benefit drop depending on the NN's slope, transitioning to the projected red line using after March 2019. The green line represents the actual net benefit drop for the XGBoost model up to March 2019, changing to the projected purple line after March 2019. A clinical significant decrease (0.9035 to 0.8808) is shown for NN but not for Xgboost (0.9051 to 0.8962).*
>
> *Figure 5. The actual and projected net benefit drift for NN and XGBoost models over time.*
>
> *Discussion*
>
> *However, through projecting the net benefit into the year 2030 based on the fitted linear regression, the decreases in the net benefit for Xgboost over time was shown to be clinically insignificant. On the contrary, the Neural network model showed a clinically significant drop in net benefit.*

The above important points have been incorporated in lines 372-375, 548-557, and 641-644.

Thank you again for your expert and invaluable review; this has been really thought provoking and has made significant contributions to improving the journal's quality of output. Thanks.

We hope this now addresses all your queries. Thank you.

References

1. Anonymous. Peer review of "Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis". JMIRx Med 2024;5:e60428. [doi: 10.2196/60428]

2. Dong T, Sinha S, Zhai B, et al. Performance drift in machine learning models for cardiac surgery risk prediction: retrospective analysis. JMIRx Med 2024;5:e45973. [doi: 10.2196/45973]
3. Zeng J. Peer review of “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis". JMIRx Med 2024;5:e60280. [doi: 10.2196/60280]
4. Kang X. The effect of color on short-term memory in information visualization. In: VINCI ’16: Proceedings of the 9th International Symposium on Visual Information Communication and Interaction: Association for Computing Machinery; 2016:144-145. [doi: 10.1145/2968220.2968237]
5. Dong T, Sinha S, Zhai B, et al. Cardiac surgery risk prediction using ensemble machine learning to incorporate legacy risk scores: a benchmarking study. Digit Health 2023 Jul 20;9:20552076231187605. [doi: 10.1177/20552076231187605] [Medline: 37492033]
6. Dong T, Sinha S, Fudulu DP, et al. Random effects adjustment in machine learning models for cardiac surgery risk prediction: a benchmarking study. medRxiv. Preprint posted online on Jun 12, 2023. [doi: 10.2196/preprints.49914]
7. Sinha S, Dong T, Dimagli A, et al. Comparison of machine learning techniques in prediction of mortality following cardiac surgery: analysis of over 220 000 patients from a large national database. Eur J Cardiothorac Surg 2023 Jun 1;63(6):ezad183. [doi: 10.1093/ejcts/ezad183] [Medline: 37154705]
8. Hickey GL, Blackstone EH. External model validation of binary clinical risk prediction models in cardiovascular and thoracic surgery. J Thorac Cardiovasc Surg 2016 Aug;152(2):351-355. [doi: 10.1016/j.jtcvs.2016.04.023] [Medline: 27215928]
9. Benedetto U, Sinha S, Lyon M, et al. Can machine learning improve mortality prediction following cardiac surgery? Eur J Cardiothorac Surg 2020 Dec 1;58(6):1130-1136. [doi: 10.1093/ejcts/ezaa229] [Medline: 32810233]
10. Zeng J, Zhang D, Lin S, et al. Comparative analysis of machine learning vs. traditional modeling approaches for predicting in-hospital mortality after cardiac surgery: temporal and spatial external validation based on a nationwide cardiac surgery registry. Eur Heart J Qual Care Clin Outcomes 2024 Mar 1;10(2):121-131. [doi: 10.1093/ehjqcco/qcad028] [Medline: 37218710]
11. Mori M, Durant TJS, Huang C, et al. Toward dynamic risk prediction of outcomes after coronary artery bypass graft: improving risk prediction with intraoperative events using gradient boosting. Circ Cardiovasc Qual Outcomes 2021 Jun;14(6):e007363. [doi: 10.1161/CIRCOUTCOMES.120.007363] [Medline: 34078100]

Abbreviations

AUC: area under the curve
CEM: clinical effectiveness metric
ECE: expected calibration error
EuroSCORE: European System for Cardiac Operative Risk Evaluation
NACSA: National Adult Cardiac Surgery Audit
SVM: support vector machine
XGBoost: extreme gradient boosting

Edited by E Meinert; submitted 09.05.24; this is a non–peer-reviewed article;accepted 09.05.24; published 12.06.24.

Please cite as:
Dong T, Sinha S, Zhai B, Fudulu D, Chan J, Narayan P, Judge A, Caputo M, Dimagli A, Benedetto U, Angelini GD
Authors’ Response to Peer Reviews of “Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis”
JMIRx Med 2024;5:e60384
URL: https://xmed.jmir.org/2024/1/e60384
doi:10.2196/60384

Authors' Response to Peer Reviews of "Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study"

Zaeem Ul Haq[1,*], PhD; Ayesha Naeem[2,*], MA; Durayya Zaeem[3], BSc; Mohina Sohail[4], MD; Noor ul Ain Pervaiz[4], MA

1

2

3

4

*these authors contributed equally

Corresponding Author:
Zaeem Ul Haq, PhD

Related Articles:

Companion article: https://preprints.jmir.org/preprint/48213

Companion article: https://med.jmirx.org/2024/1/e60430

Companion article: https://med.jmirx.org/2024/1/e60429

Companion article: https://med.jmirx.org/2024/1/e48213

(JMIRx Med 2024;5:e60266) doi:10.2196/60266

KEYWORDS

primary health care; mother and child health; community health worker; slums; digital applications; health communication.

This is the authors' response to peer-review reports for "Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study."

Round 1 Review

Reviewer BM [1]

General Comments

This study [2] draws on multiple sources of data to assess the efficacy and feasibility of a video-based mobile health (mHealth) tablet intervention used to train and equip 10 community health workers (CHWs) in two slums in Pakistan. The overall strength of the paper is that the authors have collected in-depth qualitative data that can help inform the field on how to build and distribute such an intervention to improve the needs in low-resource communities. The paper should be strengthened by pinpointing the unique and new contributions of the study findings to inform the field on digital health education interventions for CHWs. While the intervention is described in detail, more work is needed to explicate why this study expands our understanding of digital health education for CHWs in deprived settings.

Response: Thank you very much for appreciating the strengths of our study and helping us identify the weaknesses. In the revised version, we have reframed our paper in the context of communication capacity building of the CHWs serving within the primary health care system in low- and middle-income countries. We explicate with the help of data from the digital form of IEC materials, which are more interesting to the audience, improves fidelity during the implementation, and is not very costly.

Specific Comments

Major Comments

1. What the authors have found is largely expected and demonstrated in other work globally. Not surprisingly, they find that there is a severe lack of knowledge and critical need for education among CHWs and their patients to bring about behaviors that can improve maternal child health in low-resource communities. It is also not surprising that a well-resourced, highly supported, small-scale pilot (of just 10 CHWs trained) would be successful. While these findings are

all important to describe in detail (as the authors have done), to provide more value in this field of work, I suggest the paper more explicitly emphasize what contribution the study adds to the literature. What are the new and important takeaways to improve how mHealth education interventions for CHWs are developed?

Response: We appreciate this comment as it helped us position our findings in a better way. We realize that a small sample of health workers would show better results when they receive focused training (Discussion section, page 12, lines 18 - 20). In the revised version we highlight that our CHWs did not have prior training or experience of working in the health sector. So, their absorption of the knowledge and fidelity to the intervention is meaningful (Discussion section, page 11, line 48 onwards). We also highlight that using behavior change theory provides a structure to the home visits and keeps the health worker and their audience attentive (page 11, lines 20 - 27), while the videos ensure consistency of the content across all health providers. That we were able to produce the videos and the digital app on a low budget is also an important takeaway that we highlight in the Methods (page 6, lines 26 and 27) and Discussion (page 11, lines 39 - 42) sections.

2. In addition to more explicitly pointing to the contributions of their findings, the paper could be strengthened considerably by a discussion of how their findings can inform how this small-scale pilot can be taken to scale effectively. The authors allude to this, but I think more could be added with regard to cost-effectiveness. It sounds like an expensive and involved intervention—to my understanding, providing a tablet to CHWs, hosting a two-day training, overseeing an apprentice week, and a refresher training, all on top of development of 14 videos and a calendar for patients. Information on the costs to develop and implement this intervention could be better described, and I would appreciate a critical lens on what would be needed to scale, including identification of barriers. I think the limitations they described with respect to CHW availability, connectivity, etc, should be folded into this discussion. I think this would set up well the ongoing work they describe to test real-world effectiveness in 250 mother-infant pairs.

Response: Thank you for this helpful comment; it made us realize that we need to include more information on costs, and what it really means for Sehat Ghar. The development of the videos in the application was not very costly in our case, and we have included this information in the intervention description (page 6, lines 26 and 27) and discussed its implications in the Discussion section (page 11, line 39). In fact, our observation during this study and afterward is that with most smartphones, recording and editing videos are not a problem, and a number of young creative professionals are available who can do this on a low budget. Being primarily a health education intervention, our application did not need installations of updates or uploading of data from the field, which minimized its dependence on high-speed internet. We have discussed these advantages as well as the challenges to upscale in the Discussion section (page 12, lines 25 - 32).

3. A corollary to the above comment, because the intervention involves so many moving parts (ie, provision of a device, development of videos, in-person training and supervision), do their findings point to particular components of the intervention that are particularly important?

Response: Thanks for this insight. Sehat Ghar is a "whole" that is comprised of many "parts," and it is difficult to tease out a few to be the most valuable. Moreover, we were not looking for critical elements of our intervention during the formative phase and lack data to make a substantive statement. However, in the last two paragraphs of the Discussion section (page 12), we have indicated that something can be initiated through community members in areas that are totally ignored by the public sector. Likewise, videos might be one element that could be disseminated via several media.

4. The introduction of the paper starts by highlighting the distinctions between the definitions of inequality, disparity, and inequity. I don't think the comparison adds any value to the introduction. In fact, I was confused because, after the first paragraph, implications of an expensive mHealth intervention for equity are not discussed at all. Just because a study is conducted in a low-resource setting, does not mean that it is working to improve equity. If the authors want to focus on equity, I would appreciate a more critical lens on how the high costs of a digital intervention met with barriers like internet connectivity improve the situation of the poorest communities. (Otherwise, I suggest changing the introduction paragraph.) A video-based tablet intervention that relies upon internet could ostensibly be more effective in communities with more infrastructure and resources, and when scaled more widely in better resourced communities, digital interventions may actually broaden the gap between the haves and the have-nots. How can we think about ways digital interventions can be implemented to ensure this does not happen? (Is this why the in-person CHW-to-patient link is so important? Can this be unpacked?)

Response: This is a very valuable insight, which made us rethink our current paper. We realize that we are working with a marginalized community, and our work will contribute to not only improving their behaviors in the household but also helping them get better engagement from the health system. However, this can make sense when we have the data from the rollout of the intervention to the 250 families. For now, therefore, we have removed the context of inequity and situated our work in the context of the lack of primary health care services and the communication capacity building of health workers (Introduction, page 3).

5. The phase two findings draw heavily upon the "qualitative feedback" obtained from CHWs and mothers about the Sehat Ghar application and tablet use, but there are scant details on how these data were collected and analyzed. If these qualitative findings are so prominent in the results and not merely anecdotal and complementary to other findings, they need to be described with the level of detail the preintervention in-depth interviews and focus group discussions were described. Were semistructured guides used? What framework was applied to the development of qualitative protocols? How were the data coded and analyzed? How many CHWs and mothers participated?

Response: Thank you for this comment. Indeed, the data collection and analysis needed more organization and clarity because of the several phases involved in this formative study. In the revised submission, we have included an overarching question (page 5, lines 14-16) that served as a framework. We have included more details about data collection tools and methods along with clarifying qualitative and quantitative analysis for all three phases of the study (page 5, sections on data collection and data management and analysis).

6. How were the two slums selected for focus? What inclusion/exclusion criteria, if any, were applied when thinking about geographic selection? How do the two slums generalize to the larger set of slums in Islamabad?

Response: Thanks for indicating this requirement, which we have fulfilled by including the selection criteria in the Methods section. At this stage, we were focused on improving health behaviors at the household level including the improvement in health seeking for mothers and children from the formal health system. So, we had to select clusters close to the public sector health facility. Including some financial support would be required for traveling for those slums that are far; however, providing transportation was not within the remit of this project. We aim to include financial support in the next iterations of our study.

7. More of a description of the health systems in the slums is helpful for readers who are not familiar. How do CHWs fit into the larger health system? To my understanding, the 15 CHWs that were included in this study were completely new to the profession as "we identified volunteer women willing to become CHWs." Why focus on completely new volunteers for the study rather than drawing upon the existing CHW workforce? Is it because there were not CHWs already at work in these areas? If there are other CHWs already serving these slums, can this be better described? Please also speak to the generalizability of the findings given that the CHWs in this study started with a much lower lack of knowledge given their novice status. If the study were to be done with experienced CHWs, perhaps the delta in knowledge gains would not be nearly as large.

Response: These are very good points; thank you. The lady health workers of Pakistan's national program primarily work in rural areas. Slums are ignored, and these two from Islamabad also belong to the same category. We describe this lack of primary health coverage in the first part of the Methods section (page 4, lines 2-8). We also agree that the large gradient of knowledge improvement that we achieved may have been due to the novice CHWs. We acknowledge this limitation and some of the challenges that this intervention may face with the lady health workers of the national program in the Discussion section (page 12, lines 16-33).

8. A more detailed description of how participants (health care providers, CHWs, and mothers) were recruited for the study is needed. What was the sampling frame? What were the inclusion/exclusion criteria? What was the consent rate? What roles did the health care providers hold (ie, were they doctors, nurses, other roles)?

Response: We have made information about all these categories more explicit in the Study Participants section (page 5, lines 2-13).

9. What were the protocols for conducting observations? Were they unannounced or were CHWs prepared in advance to know that the supervisors would be conducting the observations? Can you address limitations with respect to bias, as individuals generally behave quite differently when observed?

Response: Thanks for this question. We have included the relevant information in the Data Collection section (page 5, lines 23-29) and discussed the limitations (page 12, lines 21-29), as these are important.

10. Do the authors have any analytics (ie, frequency of video views, engagement with the app) from the tablet/application that can be used to support the observation data and qualitative feedback on the intervention feasibility?

Response: Thanks for this question. Using digital analytics was not within the project scope and hence it was not considered during the formative phase.

Minor Comments

1. In the Abstract, identify the larger geographic location of the communities.

Response: Thanks. We have included the geographical location in the first paragraph of the Abstract.

2. Is there a more recent citation than the 2015 reference used for [3]?

Response: We have replaced this with the most recent reference from Census 2023 of Pakistan [4].

3. The Methods section says that the initial five transcripts were coded independently by two members of the team. What about the remaining transcripts? Were there any checks/reconciliation on the coding of the remaining transcripts?

Response: The initial transcripts were used to develop a code list, which the two researchers discussed and finalized. The final code list was used for analyzing all transcripts (page 5, lines 37-41).

4. Consider moving some of the details of the intervention, including on page 7 of the Results, to the Methods section. When reading the Methods, I expected to see more of these details there and am a bit confused as to why they are included in the Results.

Response: Thanks for this observation. It helped us in reorganizing the Methods and Results sections. In the revised version, we have moved the intervention details to the Methods section with an independent subheading "The Intervention" (page 6).

5. Suggest not paraphrasing Steve Jobs in the Discussion section.

Response: Thanks, we have deleted that part in the revised submission.

6. The manuscript states that this pilot was conducted in 2018. The study also notes that ongoing work with 250 mother-infant pairs is currently being conducted, now 5 years later. Given how much has happened in the world, I am curious if the authors have any reflections on how the pandemic has changed the way we should understand and reflect their findings. (The pandemic need not be addressed in the manuscript, but the second to last paragraph of the Discussion talks about health emergencies. I am skeptical how such an involved pilot could be so quickly mobilized to respond to health emergencies. The authors should reflect on this if they believe findings point to this as a possibility. I also think the detailed statistics about flooding in Pakistan and other emergencies are out of scope for the paper. I don't believe anyone needs convincing that health emergencies of this nature exist.)

Response: Thanks for this valuable input. We have deleted the content about emergencies and floods, and included our reflections on how this study might contribute to response efforts in outbreak situations (page 12, lines 11-19).

Anonymous [5]

General Comments

This paper addresses a principal issue, especially for the developing world where the valuable lives of mothers and children can easily be prevented. However, of course, a big challenge in the proposed solution is the availability of Android devices that are also connected to the internet. This is a limitation, therefore, to be added. Another area that needs to be addressed is related to cultural acceptance and sensitivity to using technology, particularly during the prenatal stage. Also, I noted that the diagrams are not clearly developed and placed in the appropriate places. I am happy with the qualitative part but unfortunately not an authority on quantitative; thus, this part should be vetted by a quantitative expert.

Response: Thanks very much for appreciating the value this study brings to the lives of mothers and children living in underserved areas.

- In the revised version, we have mentioned the costs of development (which were not much) and the occasional need for internet, as data were not required to upload or download while in the field.
- In our discussions with participant mothers and the community, we did not come across any myths or apprehensions about using digital technology while being pregnant.
- We have revised our diagram and placed it in the appropriate place, as advised.

Specific Comments

Major Comments

1. This is an excellent topic requiring continuous literature development.

Response: Thank you for appreciating our contribution.

2. The method used is mixed, whereas I would have preferred the total qualitative inquiry considering the set aims and objectives.

Response: Thanks for this observation. We relied mostly on qualitative methods and have used quantitative only where it could improve the robustness of the study.

3. The authors need to pay attention to the sociocultural realities of the context; therefore, either address them or acknowledge them as limitations.

Response: Thanks for this important observation. Our biggest constraining reality was the absence of a formally engaged CHW from the public sector, which we circumvented by engaging volunteers. Second, we did not propose biomedical steps (eg, frequent ultrasound or going to health facility for each and every problem) that would entail high costs for the family. Rather, we empowered them with knowledge and skills to identify and troubleshoot problems that could be addressed at home and have a new understanding that something can be done even in the worst circumstances. These are small bits of value, expected to help women and their families living in an underserved context.

Minor Comments

4. The diagrams need to be appropriately designed and placed in the paper.

Response: Thanks for this input. We have reduced our diagrams to one and have placed it within the text (page 4) as advised.

Round 2 Review

Reviewer BM

The authors have very thoughtfully and substantially revised the paper, making clear the methods and contributions of the study. I appreciate the detail with which the authors pointed to their edits in the revised manuscript and am satisfied with their changes.

Response: Thank you very much for appreciating the effort we made in revising the draft. We were able to considerably improve it because of the extensive input you gave—thanks again!

At this point, I suggest only very minor revisions, asking authors to check grammar and conduct a copyedit of the paper. There are instances where a careful copyedit will improve the overall reading experience of the paper. For instance, in the Abstract, I suggest the following changes:

1. *Drop the "the" in "Can the information-technology (IT) help these CHWs?"*
2. *Add a comma after "application" in "We explored answers through development and feasibility testing of Sehat Ghar, an android-based digital application to improve the communication capacity of volunteer CHWs in two slums of Islamabad."*
3. *Do not capitalize "Focus Group Discussions" in the Methods section.*

Response: In light of your overarching comment about grammar and copyediting, we have revised the entire manuscript including the three specific observations you made in the Abstract:

1. We have dropped "the" in line 6 of the Abstract.

2. Added a comma after “application” in line 8 of the Abstract.
3. Dropped capitalization of “Focus Group Discussions” in line 13 of the Abstract.

We hope that our effort in improving the language and grammar has resulted in a better expression, overall, and it meets the required standards. We are highly indebted to your kind time and valuable inputs and wish to close our response with a sincere thanks once again.

Kind regards.

References

1. Johnston JS. Peer review of “Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study". JMIRx Med 2024;5:e60430. [doi: 10.2196/60430]
2. Haq ZU, Naeem A, Zaeem D, Sohail M, Pervaiz NUA. Development of a digital platform to promote mother and child health in underserved areas of a lower-middle-income country: mixed methods formative study. JMIR Med 2024;5:e48213. [doi: 10.2196/48213]
3. Government of Pakistan. National Report of Pakistan for Habitat III. Habitat III. 2015 Jun. URL: https://habitat3.org/wp-content/uploads/Pakistan-Final-in-English.pdf [accessed 2024-06-24]
4. Announcement of results of 7th population and housing census-2023 'The Digital Census'. Pakistan Bureau of Statistics. URL: https://www.pbs.gov.pk/content/announcement-results-7th-population-and-housing-census-2023-digitalcensus [accessed 2024-06-21]
5. Anonymous. Peer review of “Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study". JMIRx Med 2024;5:e60429. [doi: 10.2196/60429]

Abbreviations

CHW: community health worker
mHealth: mobile health

Edited by E Meinert; submitted 06.05.24; this is a non–peer-reviewed article;accepted 06.05.24; published 31.07.24.

Please cite as:
Haq ZU, Naeem A, Zaeem D, Sohail M, Pervaiz NUA
Authors’ Response to Peer Reviews of “Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study”
JMIRx Med 2024;5:e60266
URL: https://xmed.jmir.org/2024/1/e60266
doi:10.2196/60266

Authors' Response to Peer Reviews of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"

Troy B Puga[1,*], BSc; Joshua Schafer[2,*], BSc; Grace Thiel[1,*], BSc; Nicholas Scigliano[1,*], BSc; Tiffany Ruan[1,*], BSc; Andres Toledo[1,*], BSc; Prince N Agbedanu[3,*], BSc, PhD; Kevin Treffer[4,*], BSc, DO

1

2

3

4

[*]all authors contributed equally

Corresponding Author:
Troy B Puga, BSc

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45688

Companion article: https://med.jmirx.org/2024/1/e55900

Companion article: https://med.jmirx.org/2024/1/e55997

Companion article: https://med.jmirx.org/2024/1/e56040

Companion article: https://med.jmirx.org/2024/1/e56169

Companion article: https://med.jmirx.org/2024/1/e45688

(JMIRx Med 2024;5:e55863) doi:10.2196/55863

KEYWORDS

COVID-19; injury; prevalence; adaptation; sports medicine; follow-up; training; football; epidemiology; sport; athlete; athletic; injuries

This is the authors' response to peer-review reports for "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study."

Round 1 Review

Dear authors, the peer review has raised a significant challenge to your statistical analysis by 1 of the reviewers. This feedback will need to be carefully addressed and/or rebutted in order for your paper to proceed at JMIRx Med.

Response: We believe that our manuscript [1] follows the appropriate format. In addition, we believe that we have addressed the statistical analysis. We also sought outside assistance to help ensure that our biostatistics methods are sound. We also would be willing to accept a transfer to *JMIR Data* if our paper is not accepted in *JMIRx* and a fee waiver can be provided.

In addition, our research team would like to highlight the derogatory, rude, and offensive review by Reviewer AC. While we appreciate the criticisms of our paper and want peer reviewers to remain vigilant in upholding the highest standards, reviewer AC used unprofessional and derogatory language toward our research team with comments such as "their precious paper." There is no place in academia for derogatory and offensive language, which is part of the guidelines in the "(for reviewers) How to write a high-quality peer review" article on the JMIR website [2].

While we accept the criticisms of our paper, we do not accept the offensive and hostile language that Reviewer AC provided to our research team. We thank you for your concerns and hope this feedback will help guide your peer-review processes in the future.

Reviewer O [3]

General Comments

I would like to commend the authors on performing a follow-up study. This well-executed study provides a good comparison with how COVID-19 disrupted global schedules and the impact it had on the sports sector. The previous version of this study identified how a lack of training greatly increased chances of injuries in National Football League (NFL) athletes. In the

current study, the authors have identified how a timely training program can reduce the injury time of athletes. A few points require clarification, such as:

1. What is the starting month of the NFL season? The COVID-19 lockdown was implemented from late March 2020 to late May 2020 (mentioned in the current study). Did this fall at the start, at the middle, or toward the end of the training phase of the athletes?

Response: This fell at the start and through most of the training period. While the lockdown ended in May, most players did not return to in-person activities until the end of July. Preseason games were eliminated and practices in training camp were far fewer in number, with only a 20-day acclimation period [4]. While the official training camp typically only lasts around 2 weeks in the second part of July, minicamps and other unofficial training were eliminated.

The following was added: "While facilities opened back up in May, players were unable to return until the end of July [5]. This shutdown eliminated most of the training period, giving players a narrow 20-day period to reacclimate before the start of the 2020 season [5]."

2. Were athletes provided any equipment at home, or were they recommended any training protocol by their team's coaches (the way some clubs in the English Premier League provided gym equipment to their players at home to maintain fitness, or had online practice sessions with their players)?

Response: Teams were allowed to host virtual instructions and online workouts. They could provide players with up to US $1500 worth of at-home workout equipment [6]. This was optional for players, and they still lacked access to trainers and other department staff and facilities that the NFL is known for having. It can be argued that even at-home workouts are not to the higher standard of facility-based workouts and lack the careful attention and recovery treatments.

The following was added: "It is important to note that although teams could provide players with up to US $1500 worth of at-home training equipment, the players still lacked access to on-site athletic trainers and recovery facilities and were unable to partake in normal preseason training [6]. Any at-home workouts were considered voluntary [6]."

3. When lockdown restrictions were lifted, how much time were the athletes provided to restore their match fitness?

Response: About 20 days were provided.

4. How long does it take for detraining to set in, and how quickly can athletes regain their lost fitness?

Response: The literature is varied on how long it takes for detraining to set in. Further, detraining can occur at various times when different physiologic changes are considered, such as a change in VO_{2max} and muscle components like mitochondrial proteins. Retraining is similarly widely variable, corresponding to the athlete, pretraining level, time off, and the level of detraining [5]. Due to the wide variability, future research is necessary to determine the exact length of time required for detraining to set in.

We did include discussion of the variability of time to regain lost fitness in the introduction.

Minor Comments

5. Change the tense of some sentences. In the introduction, use "We hypothesized" instead of "we hypothesize" and change the next line to "injury prevalence for the 2021 and 2022 seasons WOULD be lower than the 2020 NFL season..."

Response: We believe that we have corrected all tense errors.

Anonymous [7]

This paper provides epidemiological data on injury incidences in the NFL before and after the COVID-19 lockdown. This paper has the potential to be clinically meaningful; however, it has major flaws that should be addressed before it is reconsidered.

Specific Comments

Major Comments

1. The authors state that "This is the first large scale opportunity to demonstrate the effects of these principles and how they are important to understanding injury epidemiology." However, there are studies that have looked at the effect of COVID-19 on other sporting leagues, for example, Bundesliga, Major League Soccer (MLS), etc. All this has been published and should be cited.

Response: The Bundesliga study [8] did not include analysis for the years following the pandemic. Another study [9] looked at injury incidence as related to COVID-19 infection. One study [10] looked at a 41-day period of injury epidemiology for 3 periods, 2 of which were after the lockdown but still within the 2020 time period. A study focused on the National Basketball Association (NBA) [11], but again, it only considered injuries for a year following the pandemic lockdowns.

We have changed this sentence in the abstract to address these concerns: "This is the first large-scale and long-term opportunity to demonstrate the effects of these principles and how they are important to understanding injury epidemiology."

2. Can the authors please confirm or comment on the potential accuracy of this open data? What validity checks were employed to demonstrate that these data are accurate?

Response: NFL teams are required to report injuries to players throughout the week [12]. Each team reports injuries weekly, to which individual websites were referenced. The NFL website was used in the event a team report was not available in order to resolve any discrepancies.

The following was added: "Teams are required to report injuries throughout the week of a game, so any and all information should be considered accurate [12]."

3. How can the authors be sure that the increase in injuries was due to COVID-19?

Response: We cannot be certain. However, by including a comparison of 2 years prior to the COVID-19 pandemic and 2 years after, we would hope that this would account for any additional changes. We have added a sentence in the limitations

to directly address this: "It is also possible that other factors influenced the findings of this study. However, the span of years included would hopefully account for any other possible differences besides the COVID-19–induced lockdown."

4. What is meant by injuries? Soft tissue injuries or concussions? Perhaps an analysis on time-loss injuries would be more beneficial and add value

Response: In our study design methodology, we define injuries as "Contact injuries were included in this study as this is a nonmodifiable risk factor that cannot be controlled due to football being a contact sport [13,14]. COVID-19 infection, sick days, and nonmedical days off were not included in the injury tally. Illnesses were not included in this study, because illnesses are not considered a physical injury and should be reported separately from injuries when performing injury epidemiological studies [13,15]." This includes any soft tissue injuries and concussions. While it might be of interest to assess time-loss injuries, this invites the debate on whether to assess time to return to activity or time to return to full play. Only time to return to activity could be assessed from the publicly available data, and some injuries listed do not fully result in the loss of participation [12]. Further, this was not something considered in the previous study, so we would lack the ability to compare the current findings to the previous study, which was part of the goal of this project.

The following was added: "All other soft tissue injuries and concussions were included during the first week of the associated injury report."

Anonymous [16]

This paper, "COVID-19 NFL Injury Prevalence Analysis, A Follow-Up Study," is an interesting read. The conclusions drawn in the paper are supported by the data. All the related works are cited appropriately. The limitations of the presented study are discussed appropriately.

My only concern is about the 2 histograms. The authors should make these histograms more clear, readable, and engaging. Use standard deviation or error bars wherever applicable.

Response: Thank you for your comments; we have added error bars to the figures. We have also changed Figure 2 to the outcome measure to make it more appealing.

Reviewer AC [17]

General Comments

My Review—COVID-19 NFL Injury Prevalence Analysis, A Follow-Up Study

Throughout the manuscript, including in the title, authors say they analyzed the prevalence of injuries. This is incorrect. They have not analyzed injury prevalence; they did not even collect the data required for such an analysis. Instead, they collected injury incidence data and analyzed them.

Response: We have removed prevalence from the title and correspondingly changed it to "NFL Injury Analysis."

The primary component of this study is analyzing publicly available data to make a conclusion. I have a major concern regarding the statistical analysis the authors have performed. They have collected injury incidence data for each week for each team over the season from publicly available sources. This includes injuries from the same team for each week, which is repeated data. They then calculated the mean per week per team. They had 32 teams and therefore have 32 means for a season. They then compared the mean of those means between seasons using an unpaired t *test. First, this analysis totally ignores complications due to nonindependence in repeated data. Second, how can we understand the comparison of the means of means? Third, they compared each possible pairs of years. They ignored the multiple comparison issue. This analysis is totally inappropriate. I am not going to accept results of this analysis, or any conclusion based on these results. This is an issue that cannot be rescued by a revision.*

Response: Thank you for your comments. However, a major problem with your review is that at no point were any data repeated, as it is clearly defined in the methods that only new unique injuries were included. There was never any repeat of any data at any point during this study, and therefore, this statement is completely and entirely false. The research team is unaware how this assumption or conclusion was made, as this was clearly stated in the *Methods* section of both the first paper and this paper. Therefore, this claim and the claim of nonindependence is completely false.

The question of the means of means is due to the fact that in 2021, the NFL expanded to an 18-week season, whereas before it was only 17 weeks. While not ideal, the only way to compare to different season lengths is to standardize them by dividing by the number of weeks. This is how we produced this injury rate that we used for comparison. We have added more information into our *Methods* section to ensure that there is no question as to why this was undertaken. In addition, we included the 2018 and 2019 seasons in our new analysis, so you can see that the rate had to be conducted this way due to the fact that 2018, 2019, and 2020 seasons were 17 weeks long, and the 2021 and 2022 seasons were 18 weeks long. All of this reasoning, along with describing how data were not repeated, are clearly laid out in the *Methods* section, and we hope that you will conduct a more thorough reading of this section.

Finally, we have used a Kruskall-Wallis test with Dunn analysis to address the comparison of pairs. The comparisons are statistically significant for 2020 when compared with all the seasons. It is clear that there is significant data here. In addition, comparison of the 2019 season with the 2021 and 2022 seasons did not produce a statistically significant difference, indicating a return to normal levels. We hope that you can accept the results of this analysis. In addition, we consulted with a top leading academic university for biostatistics help in order to ensure that our methods are sound. We hope that our efforts will help you see these results as justified.

The authors say they have done similar analysis in their precious paper [13]. I now doubt the findings published there too. Unfortunately that paper was also published in JMIR. *I*

recommend that editors should consider rereviewing that paper by an independent statistical reviewer.

Response: Thank you for your comments; however, we would like to highlight that this language is offensive and derogatory and there is no place for comments like this in academia. We appreciate constructive feedback and your dedication to upholding the highest peer reviewing standards; however, we want to make note that this language is offensive and derogatory toward our research team and the previous peer reviewers. The previous paper was reviewed by peer reviewers, who believe that the original analysis was justified.

There are less severe issues as well. For example, they presented 2 figures—one is redundant in the presence of the other, because the numbers in Figure 1 divided by the number of weeks are the numbers in Figure 2. Further, none of the numbers in any of these figures are the outcome measure they used in the statistical analysis. Therefore, the usefulness of them is limited only to describing the raw data.

Response: We have retained Figure 1 as it provides the total number of injuries in the data. We have changed Figure 2 to reflect the outcome measure to make it more useful.

Even if the analysis is correct, they have a fundamental limitation in their interpretation of the results. Their conclusions are based on the underlying assumption that the observed statistical differences were driven by training opportunities. There was no justification for that assumption. How can the authors claim none of the other possible influencing factors changed?

Response: We cannot confirm or deny that other factors could have influenced these changes, and it is clearly stated in our first paper and this paper within the discussion that other factors could have played a role. However, the largest precipitating change during this time period was COVID-19, which led to limited training opportunities. This is factual evidence and does create a pathway for this epidemiological spike in 2020. It is clearly discussed in both papers within the introductions and discussions that the loss of training can induce detraining, which leads to a predisposition to injuries. This is why there is an epidemiological spike in the number of injuries during 2020 and is statistically significant when compared with the 2018, 2019, 2021, and 2022 seasons. This is also why there is no statistical significance when the 2019 season is compared with the 2021 or 2022 seasons. Therefore, our data and evidence of the events during this time support our conclusions. While other factors could have played a role, the most likely reason was due to these well-documented injuries and strength and conditioning principles.

References

1. Puga TB, Schafer J, Thiel G, et al. COVID-19 National Football League (NFL) injury analysis: follow-up study. JMIRx Med 2024;5:e45688. [doi: 10.2196/45688]
2. JMIR Editorial Team. (For reviewers) how to write a high-quality peer review. JMIR. URL: https://support.jmir.org/hc/en-us/articles/16470162812827--for-reviewers-How-to-write-a-high-quality-peer-review [accessed 2024-01-31]
3. Pereira F. Peer review of “COVID-19 National Football League (NFL) injury analysis: follow-up study". JMIRx Med 2024;5:e56169. [doi: 10.2196/56169]
4. Battista J. 2020 in review: inside the most unique season in NFL history. NFL. 2021 Feb 1. URL: https://www.nfl.com/news/sidelines/2020-in-review-inside-the-most-unique-season-in-nfl-history [accessed 2023-07-18]
5. Girardi M, Casolo A, Nuccio S, Gattoni C, Capelli C. Detraining effects prevention: a new rising challenge for athletes. Front Physiol 2020;11:588784. [doi: 10.3389/fphys.2020.588784] [Medline: 33178050]
6. Battista J. Virtual offseason program set to begin April 20.: NFL; 2020 Apr 13. URL: https://www.nfl.com/news/virtual-offseason-program-set-to-begin-april-20-0ap3000001109298 [accessed 2023-07-18]
7. Anonymous. Peer review of “COVID-19 National Football League (NFL) injury analysis: follow-up study". JMIRx Med 2024;5:e56040. [doi: 10.2196/56040]
8. Krutsch W, Hadji A, Tröß T, et al. No increased injury incidence in the German Bundesliga after the SARS-Cov-2 virus Lockdown. Arch Orthop Trauma Surg 2022 Jul;142(7):1571-1578 [FREE Full text] [doi: 10.1007/s00402-021-04060-2] [Medline: 34286365]
9. Maestro A, Varillas-Delgado D, Morencos E, et al. Injury incidence increases after COVID-19 infection: a case study with a male professional football team. Int J Environ Res Public Health 2022 Aug 18;19(16):10267. [doi: 10.3390/ijerph191610267] [Medline: 36011901]
10. Mazza D, Annibaldi A, Princi G, et al. Injuries during return to sport after the COVID-19 Lockdown: an epidemiologic study of Italian professional soccer players. Orthop J Sports Med 2022 Jun;10(6):23259671221101612 [FREE Full text] [doi: 10.1177/23259671221101612] [Medline: 35722177]
11. Torres-Ronda L, Gámez I, Robertson S, Fernández J. Epidemiology and injury trends in the National Basketball Association: Pre- and per-COVID-19 (2017–2021). PLOS ONE 2022;17(2):e0263354 [FREE Full text] [doi: 10.1371/journal.pone.0263354] [Medline: 35143536]
12. NFL. 2017 personel (injury) report policy. 2017. URL: https://operations.nfl.com/media/2683/2017-nfl-injury-report-policy.pdf [accessed 2024-02-02]
13. Puga TB, Schafer J, Agbedanu PN, Treffer K. COVID-19 return to sport: NFL injury prevalence analysis. JMIRx Med 2022 Apr;3(2):e35862. [doi: 10.2196/35862] [Medline: 35511457]

14. Emery CA. Injury prevention and future research. Med Sport Sci 2005;48:179-200. [doi: 10.1159/000084289] [Medline: 16247258]
15. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Clin J Sport Med 2006 Mar;16(2):97-106. [doi: 10.1097/00042752-200603000-00003] [Medline: 16603877]
16. Anonymous. Peer review of "COVID-19 National Football League (NFL) injury analysis: follow-up study". JMIRx Med 2024;5:e55900. [doi: 10.2196/55900]
17. Samaranayaka A. Peer review of "COVID-19 National Football League (NFL) injury analysis: follow-up study". JMIRx Med 2024;5:e55997. [doi: 10.2196/55997]

Abbreviations

MLS: Major League Soccer
NBA: National Basketball Association
NFL: National Football League

Edited by E Meinert; submitted 27.12.23; this is a non–peer-reviewed article;accepted 27.12.23; published 13.02.24.

Please cite as:
Puga TB, Schafer J, Thiel G, Scigliano N, Ruan T, Toledo A, Agbedanu PN, Treffer K
Authors' Response to Peer Reviews of "COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study"
JMIRx Med 2024;5:e55863
URL: https://xmed.jmir.org/2024/1/e55863
doi:10.2196/55863

Authors' Response to Peer Reviews of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers"

Jaime Perales-Puchalt[1], MPH, PhD; Jill Peltzer[2], PhD, APRN-CNS; Monica Fracachan-Cabrera[2], MD; G Adriana Perez[3], PhD, ANP-BC; Mariana Ramírez[2], LCSW; K Allen Greiner[2], MD; Jeffrey Murray Burns[1], MD

1

2

3

Corresponding Author:
Jaime Perales-Puchalt, MPH, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/42211

Companion article: https://www.medrxiv.org/content/10.1101/2022.05.25.22275517v2

Companion article: https://med.jmirx.org/2024/1/e56444

Companion article: https://med.jmirx.org/2024/1/e56443

Companion article: https://med.jmirx.org/2024/1/e57160

Companion article: https://med.jmirx.org/2024/1/e42211

(JMIRx Med 2024;5:e56439) doi:10.2196/56439

KEYWORDS

Latino; dementia; caregiving; COVID-19; Alzheimer's disease; health disparity; qualitative research; transcript analysis; health inequality; minority population; epidemiology; peer support; social services; health care; Alzheimer's; minority; qualitative; interview; caregiver; primary care; impact; resilience; disparity; outcome; Alzheimer disease; Alzheimer

This is the authors' response to peer-review reports for "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers."

Round 1 Review

Dear Prof. Meinert,

Re: Manuscript ID #42211 entitled "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers" [1].

Enclosed is a file containing the revised manuscript by Perales-Puchalt et al [1]. Thank you for this opportunity to resubmit the present paper.

We also thank the reviewers very much for their suggestions and comments to our manuscript. We have taken all of them into consideration for producing the new version. In the following sections, we have first presented the reviewers' comments, followed by our responses.

We hope that these responses and the new manuscript prove satisfactory to you and your reviewers.

We are looking forward to hearing from you. Thank you again for your attention to our work.

Sincerely,

Jaime Perales-Puchalt on behalf of the authors

Fairway, KS; October 12, 2023

Anonymous [2]

General Comments

This paper describes a qualitative study in which the authors seek to understand the experiences of Latino families managing Alzheimer disease and related dementias (ADRD). The authors interviewed both family caregivers and primary care providers

(PCPs). This is a well-written manuscript that focuses on caregiving during COVID-19, a relatively understudied area. The authors note that this study represents secondary analyses of a larger study that focused on improving ADRD care services in primary care across settings.

Specific Comments

Major Comments

1. My key methodological critique is that the authors should follow the COREQ (Consolidated Criteria for Reporting Qualitative Research) recommendations in reporting this study. Since this is a completed study, it is possible that the authors will not meet all the criteria; however, it is still important to know which recommendations were followed and which were not.

Response: Thank you for the resource! We have edited the manual to follow these guidelines.

2. My key conceptual critique is that the authors do not provide any rationale as to why family caregivers' and PCPs' perspectives are included together. While I certainly understand that these are secondary analyses, the study introduction still needs to justify why these 2 stakeholder groups will provide perspectives that can be synthesized.

Response: We have clarified in the introduction section that "It is also important to listen to the perspectives of family caregivers and primary care providers (PCPs), who may provide a different point of view and allow triangulation. Given the central role of family in Latino culture [14] and PCPs in the health care system [15], their perspectives can provide privileged insight into the experiences of the person with ADRD, irrespective of their level of cognitive impairment, their family, and their interaction with the health care system."

3. Relatedly, the authors should provide a conceptual or theoretical framework that guided this study.

Response: We added the following to the methods section: "This study was informed by the National Institute on Minority Health and Health Disparities Research Framework, which considers the complex and multifaceted nature of minority health and health disparities [17]. This framework includes different domains of influence (biological, behavioral, physical and built environment, sociocultural environment, and health care system), as well as different levels of influence (individual, interpersonal, community, and societal) within these domains."

4. In methods, please clarify whether the main interview was 45-60 minutes and the additional COVID-19–specific questions were excluded from this time. Please also provide the interview guide as an appendix (see COREQ) and include sample questions in the manuscript.

Response: We have clarified that the 45-60 minutes included the COVID-19 questions.

5. Please clarify what is meant by "the interviewers emphasized that participants were the experts to reduce power differentials." How was this accomplished? How did the interviewer ensure that this power differential existed and that it was subsequently reduced?

Response: We have clarified this sentence, which now says: "The potentially lower theoretical ADRD expertise by family members and PCPs could create a perceived power differential with the interviewer. In fact, some family members and PCPs expressed worry about the interviewer potentially testing their ADRD knowledge. To reduce this perception of power differential, the interviewer emphasized that the interviews would ask about their experiences and that they were the experts in those experiences."

6. Consider expanding the description of theme 1 to capture the dimensions of impact noted in the subthemes, for instance, "Both caregivers and PCPs highlighted the physical, psychological and social impacts of the pandemic on patients with ADRD."

Response: We have expanded the description of theme 1 to capture the dimensions noted in the subthemes, as suggested by the reviewer.

7. The second theme does not appear to integrate the PCP and caregiver views as well as theme 1. consider splitting the current theme 2 into 2: theme 2 could be about individual coping and resilience, and theme 3 could be about systems level factors, which would include vaccination acceptance and remote communications.

Response: We have eliminated the 2 higher levels of themes and included the lower levels. In line with this comment, we have differentiated between individual supports and system supports.

8. The conclusion discusses death and formal care, which are not highlighted in the themes or results. Anchor the discussion to the results of this study.

Response: We have carefully linked the discussion with the most important results.

Minor Comments

1. Please review and correct typographical errors; for example, in data analysis, it says "fist author" instead of "first author."

Response: We have carefully proofread the manuscript.

2. The citation #16 is for focus groups, but the authors did 1:1 interviews. Please ensure that this citation is correct.

Response: Thanks for catching that one. That was an extra citation that was inserted by mistake.

3. The percentages in Table 1 are not meaningful given the small sample size. Please report only the Ns.

Response: We have deleted the percentage columns.

Reviewer FN [3]

General Comments

The authors present a compelling argument for understanding how a doubly vulnerable population in the United States (Latino persons with dementia) experienced the COVID-19 pandemic. To this end, they share the results of a thematic analysis of interviews with primary care providers and caregivers of Latino persons living with dementia. The qualitative analysis could

use better explanation and the themes could be more descriptive. Moreover, the many of themes do not capture or explain their relevance to understanding the intersectionality of dementia and Latino lives. My comments below speak to that, as well as other issues. The manuscript has a good foundation of an informative article that showcases the lived experience of this population during a critical time and could be modified to provide formative evidence for improving care, inside and outside of a pandemic.

Specific Comments

Major Comments

1. Consider making subthemes or codes more descriptive and meaningful. Good themes tell you what the story is or what the direction is at least (good or bad). Some of these are readily available (or easily modified) from sentences in the paper already (eg, "the pandemic influence[d] mental and emotional health"; "Social support was critical for reducing social isolation and its sequalae"; caregivers and persons with dementia "lost access to engaging activities during the confinement"; and "Remote communication facilitated social support"). See other qualitative research for examples. There are even good examples in the dementia care literature during the pandemic. For example:

Mitchell LL, Horn B, Stabler H, Birkeland RW, Peterson CM, Albers EA, Gaugler JE. Caring for a Relative With Dementia in Long-Term Care During the COVID-19 Pandemic: A Prospective Longitudinal Study. Innov Aging. 2023 Apr 17;7(4):igad034. doi: 10.1093/geroni/igad034. PMID: 37213326; PMCID: PMC10195573.

Harding, E., Rossi-Harries, S., Gerritzen, E.V. et al "I felt like I had been put on the shelf and forgotten about" – lasting lessons about the impact of COVID-19 on people affected by rarer dementias. BMC Geriatr 23, 392 (2023). https://doi.org/10.1186.S12877-023-03992-1

Response: We have taken this comment into account and turned the themes into descriptive statements. We have also further divided them into more specific themes, in line with the ones the reviewer mentioned.

2. Throughout: The topic is about the Latino ADRD experience, but many themes do not tap into this overlap of the Latino and ADRD experience. It is not made relevant to the ADRD experience or it is not explained how the ADRD context affected it, either in the theme itself (eg, see "poor nutrition" codes) or in the quote used to justify it (eg, see "stress" and "work" codes). A thorough review of the codes and quotes to meet this intersectionality would be beneficial.

Response: We have thoroughly revised the themes and quotes to highlight the Latino ADRD experience as much as possible. However, while most aspects refer singularly to Latino families with ADRD, some are also experienced by Latino families in general or non-Latino families with ADRD. Including these helps provide information in the discussion section on what impacts are more general to all populations and what things may be particularly salient among Latino individuals.

3. Facilitators and barriers are not often used as codes on their own but a way to categorize or further delve into aspects of other issues. What was facilitated? What was the barrier blocking? I could see weaving facilitators, barriers, and consequences into the discussions around the other codes, like informal and formal support.

Response: We have edited our codes, which no longer only state "barriers," "facilitators," or "consequences." Instead, for example, a whole theme includes the barriers to remote communication by Latino families.

4. Please clarify the methods.

a. *I have not heard of using condensed transcripts before. Why was this done? What was taken out exactly? How were "meaningful bits of text" identified?*
b. *It is not clear who was doing the coding at which times. There is the first author as a coder and then 2 additional coders, but the final sentence indicates there were just 2.*
c. *Did the first author make all the themes and do all the coding and then the other coder(s) just reviewed it? (Rather than all independently reviewing and coming together to come up with themes and then independently coding and later addressing the coding discrepancies.)*

Response: We have edited this section to further clarify how the coding was done. It now says "JP-P and MF-C independently read the interviews and notes once initially to familiarize themselves with the data, coded the content of the text by identifying codes and themes, and resolved coding disagreements through discussion and consensus."

5. Add headers in the text for each subtheme and the codes for better flow and to help readers keep track of what theme or code they are reading about.

Response: We have added headers and numbers for better flow and to help readers keep track of themes and codes.

6. For quotes in general:

a. *Consider editing any quote over 2 lines or selecting briefer quotes. Three lines is okay if compelling. If more than 3, it has to be a really good quote. There are 2 very long quotes in "other impacts."*
b. *Provide some context or active linking to the code as lead-in text.*
c. *Make sure they are absolutely relevant to the Latino and ADRD experience.*

Response: We have shortened most quotes to make them 1-2 lines long and added more context in the text.

7. For the discussion:

a. *Consider adding comparisons of findings to non-Latino ADRD COVID-19 experiences to highlight the differences this population experienced and to better showcase why continued attention to this specific population is warranted. See the Mitchell et al and Harding et al papers cited above as potential comparison points.*
b. *Why is circular migration being brought up? As written, this does not appear to be ADRD related and was only*

lightly discussed in the results (though, it was not clear if it is ADRD related in the results either).

c. *"The fact that some PCPs suspected an unknown longer-term impact of the COVID-19 pandemic warrants further longitudinal research into this topic." While true, it is strange to frame ongoing need to understand this based on participant responses alone. Also, it needs to be related to ADRD.*

d. *"Caregivers' reports on healthcare providers' confusion between ADRD and COVID-19 infection symptoms warrants research to improve diagnosis and severity assessments of both conditions." Where did this come from? It feels unsupported from the evidence provided in the current study and an odd place to end the discussion.*

Response: We have focused the discussion on similarities and differences with previous studies on the general population and deleted the discussion on circular migration, diagnosis confusion, and unknown consequences of the pandemic.

8. For the conclusion:

a. *"This pandemic has revealed many of the barriers that Latino families with ADRD face, and in most cases, this has exacerbated previous barriers. However, with every crisis comes an opportunity for improvement, which will hopefully translate into improved conditions among Latino families with ADRD." This does not really say anything; be specific regarding the barriers and what could be improved. You could succinctly use the start of the next sentence to end this one.*

b. *"These improved conditions might include more equitable access to health care and community services, a better quality of these services, subsidized formal and informal supports, and flexible hybrid means of communication." Which would lead to…or mean what for the (public) health of Latino persons living with ADRD and their carers? What is the overall takeaway pertaining to health or public health?*

c. *The discussion and conclusions could be broadened out to medical care in general if the issues appear to also be independent of the pandemic.*

Response: We have connected the conclusion to the results and discussion and have been more specific in our statements.

Minor Comments

9. Mention the United States as the target population in the abstract.

Response: We edited the abstract to specify that it was in the United States.

10. Typos: "The fist author also condensed" and "work in the meat packing plan industry."

Response: We have edited these and other typos.

11. Clarify "To make bring rigor and validity." Or is there a typo here?

Response: It is a typo, which we have removed. Thank you!

12. "To make bring rigor and validity to the research process, the interviewer used active listening techniques during the interview aimed at confirming the information shared by participants. The interviewer also emphasized the fact that participants were the experts in their experiences to reduce power differentials." This belongs in the methods, not analysis.

Response: We have moved this section above "Data analysis."

13. "[Explains how after the lockdown, the care recipient only remembers long term memories]. So, I was thinking that all the time she was locked down here because of the cold weather and COVID might have affected her more." Avoid total paraphrasing and provide (translated) direct quotes. Or put the paraphrase as context before the quote.

Response: We have put the paraphrasing as the context before quotes to avoid total paraphrasing.

14. "...PCPs had to reduce physical contact with care recipients, which reduced their chance to convey warmth to their patients." A quote here would be nice.

Response: We added a quote for this.

15. "This fear was not unfounded. Prior to the availability of the vaccine, caregivers and care recipients acquired COVID-19. As Latino older adults, they were at an increased risk for complications including death, causing significant chronic concern and fear." This is useful for the introduction (and the vulnerability was discussed) and discussion but should not be a part of the results. I suggest omitting this.

Response: We have moved that section to the discussion section.

16. "Fourth, care recipients and PCPs highlighted the frequency and severity of depressed mood among caregivers and care recipients, especially during the lockdown due to lack of social support and social isolation. The PCPs noted that lack of social support and social isolation due to lockdown negatively impacted mood, sharing:" This is redundant. Consider condensing into 1 sentence.

Response: We have reduced the redundancies by eliminating a sentence.

17. "Fourth, consequences of social support were physical, psychological, and social. Examples of physical consequences include potentially reducing mortality by providing formal and informal caregiving services. Psychological consequences include clinic and family support reducing loneliness and increasing feelings of safety. Social consequences include caregivers being allowed to accompany their care recipients during clinic visits, curbside visits allowing socialization and home care services lowering isolation." Like the barriers above this section, these feel more like they could be part of the informal or formal support codes. What are the supporting quotes? Also, it is not clear what were the consequences? What were the causes that led to the consequences?

Response: We have restructured our themes, and these are now part of the health care and community care systems themes.

18. "Third, consequences of the higher use of remote communication were both positive and negative." This should just be part of the remote communication theme description.

Response: We have restructured the themes and removed this section.

19. "...similar to other studies, the need to rely on remote communication intensified the digital divide." Reverse it—the digital divide was problematic given the need to rely on remote communication.

Response: We have edited this sentence as suggested, thank you!

20. "...for their survival" in the conclusion is a bit heavy-handed. Speak on something closer at hand in the manuscript like avoiding exposure and infection.

Reviewer GI [4]

General Comments

This paper provides an in-depth qualitative assessment of the impact and resilience factors related to the COVID-19 pandemic among Latino families with ADRD. The authors interviewed 21 family caregivers and 23 PCPs across the United States and identified 2 primary themes that characterized the experiences of the participants, involving both the impact of the pandemic and the strategies adopted to cope with the detrimental impact of the pandemic. The topic covered is of significant importance and provides important background to better understand the impact of the COVID-19 pandemic on Latino families with ADRD, with the aim of improving the quality of care and equity among the Latino community. Overall, I think that the authors should reorganize the results to better align with the aim of the study. In my specific comments I have included specific suggestions on how to make the results more organized and succinct. To strengthen the generalizability and interpretation of the findings, the authors should also include a descriptive quantitative analysis of the interviews' analysis, where the reader can examine the prevalence of each theme and subtheme among the 21 family members and 23 PCPs.

Specific Comments

Major Comments

1. The results section of the abstract is not very clear. What are the overall findings? How have the 2 themes been identified?

Response: We have mentioned the 8 themes identified in the results section of the abstract.

2. Introduction: Please outline the qualitative variables selected to investigate the aim of the study.

Response: The qualitative variable selected to investigate the aim of the study was the impact of the COVID-19 pandemic. Please find this clarification in the introduction section, last paragraph.

3. Methods:

a. *Sample and assessment: Given the qualitative nature of the study, it would be helpful if the authors included an example of a question from the interview script and also attached as an appendix the template of questions they used to direct the conversation.*
b. *Data analysis: This section is not very clear and it would help if the authors could describe the process of coding in a step-by-step manner and also make the text more succinct.*

Response: We have included an appendix with the interview guides. We have restructured the methods to make the description of coding clearer.

4. Results: The authors provide a very detailed qualitative analysis; however, more quantitative information should be provided to show the prevalence of each theme and subtheme for all the caregivers and PCPs (eg, how many PCPs reported food access and malnutrition during the COVID-19 pandemic?)

Response: This work does not use content analysis, where quantifying the data is one of the main goals. Instead, we use thematic analysis, which focuses on using themes to generate new insights about a particular phenomenon.

Minor Comments

5. In the abstract, results section, please remove the capital letters for the 2 themes and consider writing them as " Qualitative analysis of transcripts revealed two themes: (1) the impact of a global pandemic (eg, accelerated cognitive and physical decline, or caregivers choosing between risking finances and the family's infection given the work situation) and (2) developing resilience to the effects of the pandemic (eg, caregivers seeking vaccination sites, moving in with the care recipient and adopting telehealth."

Response: We have removed the capital letters from the themes and numbered them as suggested.

6. Introduction: This sentence should be revised for clarity: "As of January of 2022, Latinos represent 8% of the US 65 and older population, but 13% of COVID-19 cases in the same age group [9]." It is not clear if the 13% of cases is the percentage of COVID-19 cases in Latino individuals aged 65 years and older.

Response: Yes, it refers specifically to the 65 years and older population for both things: the US population and the COVID-19 cases. We have edited this sentence to say "As of January of 2022, among individuals aged 65 years and older, Latino individuals represent 8% of the US population but 13% of COVID-19 cases."

7. Methods:

a. *This sentence should be incorporated in the introduction or removed: "The goal of this study was to gain an in-depth understanding of the impact of the COVID-19 pandemic on Latino families with ADRD."*
b. *Who transcribed the interviews?*

Response: We deleted the sentence and edited the following sentence to express that this analysis was part of a broader study. With respect to the transcriptions, a professional team transcribed all interviews, which we mention in the methods section.

8. Results:

a. *To improve clarity, the authors should label the descriptions of the different themes as themes (1, 2, etc) and subthemes (1.1, 1.2, 1.3, etc; 2.1, 2.2, 2.3, etc). Please add this labeling both in Table 2 and in the text below to help the reader better orient into each of the themes.*

Response: This is a very good idea. Thank you! We have added these markers.

9. Page 6, last sentence: Please remove "make" from "To make bring rigor and validity."

Response: We have deleted the typo.

10. Page 14, last paragraph: The authors should edit "work" with "works."

Response: We have edited this typo. Thank you!

Round 2 Review

Dear Prof. Meinert,

Re: Manuscript ID #42211 entitled "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers."

Enclosed is a file containing the revised manuscript by Perales-Puchalt, et al [1]. Thank you for this opportunity to resubmit the present paper.

We also thank the reviewers very much for their suggestions and comments to our manuscript. We have taken all of them into consideration for producing the new version. In the following sections, we have first presented the reviewers' comments, followed by our responses.

We hope that these responses and the new manuscript prove satisfactory to you and your reviewers.

We are looking forward to hearing from you. Thank you again for your attention to our work.

Sincerely,

Jaime Perales-Puchalt on behalf of the authors

Fairway, KS; December 8, 2023

Reviewer FN

General Comments

The authors thoroughly attended to the reviewer responses. The methods are easy to understand and the rework of the themes and related quotes in the results is a great improvement. The discussion could be easier to read by breaking the large paragraphs into smaller ones. The conclusion should be revised to more specifically attend to what the study found and how it extends the literature. These and other issues are further noted:

Specific Comments

Major Comments

Major comments in order of appearance in the manuscript:

1. In the 2.2. poor nutrition theme, it is more obvious that the mailing system and financial insecurity could directly be influenced by the pandemic, but it is not clear that skills and level of impairment were affected by the pandemic. As written, it sounds more like an overarching ADRD problem rather than an ADRD issue specific to the pandemic. This should be clarified, especially since it is brought up specifically in the discussion as a unique finding.

Response: We have clarified that the more structural issues (mailing system and financial insecurity) interacted with skills and impairment, which resulted in unhealthier options being more accessible.

2. Consider reworking this sentence to clarify and streamline: "While home-delivered meals operated normally, Latino families with ADRD tried to access these for the first time during the pandemic to obtain food while reducing the risk of infection." My suggested rework is "Some Latino families with ADRD we interviewed tried to use home-delivered meals for the first time during the pandemic to reduce risk of infection."

Response: We have replaced the sentence with the alternative provided by the reviewer. Thank you for the suggestion!

3. Break large paragraphs throughout the discussion into smaller ones by more specific topic (eg, food and nutrition, work changes, and infection risk).

Response: We have separated the paragraphs by topic as suggested.

4. How exactly are fatalism and personalism related to the findings in the study? Make an explicit tie, back into the findings to make a stronger ending to this part of the discussion.

Response: We have connected fatalism and personalism more clearly to the results from our interviews.

5. Rephrase "Was this also the case for cognitive and functional decline?" into a statement rather than a question.

Response: We have replaced this question with the following statement: "Studies could explore whether this disproportionate impact also applies to cognitive and functional decline."

6. As written, the conclusion paragraph does not indicate well what the study found or how it extends the literature. The first sentence needs to be reworked—who is "their" referring to? Families were critical to "maintaining or improving" "health and quality of life," correct? Make succinct and specific mention to how the families were affected "beyond infection and physical symptoms." What were the specific barriers that were exacerbated?

Response: We have reworked the conclusions section as suggested by clarifying the ambiguous sentences and expanding on examples.

Minor Comments

1. "Other caregivers or their care recipient had been infected or were indeed infected during the interview." This sounds like the interviewer infected them. They were experiencing COVID-19 at the time of the interview?

Response: We have edited this sentence as requested. Thank you!

2. Typo in theme 4.3: "to the their."

Response: We have fixed the typo.

3. Avoid the numeric two in the quote in theme 5.3: "Mom had 2 that got COVID." Suggested rework: "Mom had two [home assistants] that got COVID."

Response: We have replaced the number with the word "two."

4. Remove the hyphen from "frequently-mentioned."

Response: We have deleted the hyphen.

Reviewer GI

General Comments

This paper provides an in-depth qualitative assessment of the impact and resilience factors related to the COVID-19 pandemic among Latino families with ADRD. The authors interviewed 21 family caregivers and 23 PCPs across the United States. They identified 2 primary themes that characterized the participants' experiences, involving both the impact of the pandemic and the strategies adopted to cope with the detrimental impact of the pandemic. The topic covered is of significant importance and provides important background to better understand the impact of the COVID-19 pandemic on Latino families with ADRD and to improve the quality of care and equity among the Latino community. The authors did a great job improving the clarity of the methods and the results. I have included other minor revisions to further improve the clarity of the text.

Specific Comments

Minor Comments

1. Results

a. *Paragraph "theme 8" (line 4): Remove "and" before "healthcare"*
b. *Paragraph 8.3 (line 3): Replace " requested" with "requesting."*

Response: We have removed the word "and" and replaced "requested" with "requesting."

2. Discussion: Please clarify the second paragraph of the "implication and future directions" section. In the first sentence, "Our findings can inform future studies. For example, participants reported pandemic-related physical and cognitive deterioration and the importance of family support." Can you clarify what type of findings can inform future research? Could you say something on the line of "Our findings regarding the physical and cognitive deterioration caused by the pandemic and the importance of family support may help inform future studies on…" In the same paragraph, there is a very minor typo; please replace "health" with "healthy."

Response: We have replaced the sentence with the one suggested by the reviewer and corrected the minor typo. Thank you!

References

1. Perales-Puchalt J, Peltzer J, Fracachan-Cabrera M, et al. Impact of the COVID-19 pandemic on Latino families with Alzheimer disease and related dementias: qualitative interviews with family caregivers and primary care providers. JMIRx Med 2024;5:e42211. [doi: 10.2196/42211]
2. Anonymous. Peer review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers". JMIRx Med 2024;5:e57160. [doi: 10.2196/57160]
3. Peterson C. Peer review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers". JMIRx Med 2024;5:e56444. [doi: 10.2196/56444]
4. Marin A. Peer review of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers". JMIRx Med 2024;5:e56443. [doi: 10.2196/56443]

Abbreviations

ADRD: Alzheimer disease and related dementias
COREQ: Consolidated Criteria for Reporting Qualitative Research
PCP: primary care providers

Edited by E Meinert; submitted 16.01.24; this is a non–peer-reviewed article;accepted 16.01.24; published 08.03.24.

Please cite as:
Perales-Puchalt J, Peltzer J, Fracachan-Cabrera M, Perez GA, Ramírez M, Greiner KA, Burns JM
Authors' Response to Peer Reviews of "Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers"
JMIRx Med 2024;5:e56439
URL: https://xmed.jmir.org/2024/1/e56439
doi:10.2196/56439

Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study

Usha Lokala[1], MS; Orchid Chetia Phukan[2], MTech; Triyasha Ghosh Dastidar[3], MSc; Francois Lamy[4], PhD; Raminta Daniulaityte[5], PhD; Amit Sheth[1], PhD

1

2

3

4

5

Corresponding Author:
Usha Lokala, MS

Related Articles:

Companion article: https://arxiv.org/abs/2304.10512

Companion article: https://med.jmirx.org/2024/1/e58317

Companion article: https://med.jmirx.org/2024/1/e58320

Companion article: https://med.jmirx.org/2024/1/e58321

Companion article: https://med.jmirx.org/2024/1/e57838

Abstract

Background: Opioid and substance misuse has become a widespread problem in the United States, leading to the "opioid crisis." The relationship between substance misuse and mental health has been extensively studied, with one possible relationship being that substance misuse causes poor mental health. However, the lack of evidence on the relationship has resulted in opioids being largely inaccessible through legal means.

Objectives: This study aims to analyze social media posts related to substance use and opioids being sold through cryptomarket listings. The study aims to use state-of-the-art deep learning models to generate sentiment and emotion from social media posts to understand users' perceptions of social media. The study also aims to investigate questions such as which synthetic opioids people are optimistic, neutral, or negative about; what kind of drugs induced fear and sorrow; what kind of drugs people love or are thankful about; which drugs people think negatively about; and which opioids cause little to no sentimental reaction.

Methods: The study used the drug abuse ontology and state-of-the-art deep learning models, including knowledge-aware Bidirectional Encoder Representations From Transformers–based models, to generate sentiment and emotion from social media posts related to substance use and opioids being sold through cryptomarket listings. The study crawled cryptomarket data and extracted posts for fentanyl, fentanyl analogs, and other novel synthetic opioids. The study performed topic analysis associated with the generated sentiments and emotions to understand which topics correlate with people's responses to various drugs. Additionally, the study analyzed time-aware neural models built on these features while considering historical sentiment and emotional activity of posts related to a drug.

Results: The study found that the most effective model performed well (statistically significant, with a macro–F_1-score of 82.12 and recall of 83.58) in identifying substance use disorder. The study also found that there were varying levels of sentiment and emotion associated with different synthetic opioids, with some drugs eliciting more positive or negative responses than others. The study identified topics that correlated with people's responses to various drugs, such as pain relief, addiction, and withdrawal symptoms.

Conclusions: The study provides insight into users' perceptions of synthetic opioids based on sentiment and emotion expressed in social media posts. The study's findings can be used to inform interventions and policies aimed at reducing substance misuse

and addressing the opioid crisis. The study demonstrates the potential of deep learning models for analyzing social media data to gain insights into public health issues.

(JMIRx Med 2024;5:e48519) doi:10.2196/48519

KEYWORDS

opioid; substance use; substance use disorder; social media; US; opioid crisis; mental health; substance misuse; crypto; dark web; users; user perception; fentanyl; synthetic opioids; United States

Introduction

Background

North America is facing the worst opioid epidemic in its history. This epidemic started with the mass diversion of pharmaceutical opioids (eg, oxycodone, hydromorphone), resulting from the strong marketing advocacy of the potential benefits of opioids [1]. The increase in opioid use disorder prevalence and pharmaceutical opioid-related overdose deaths resulted in a stricter distribution of pharmaceutical opioids, unintentionally leading to a substantial increase in heroin use among pharmaceutical opioid users [2]. The epidemic entered its third wave when novel synthetic opioids (eg, fentanyl, U-47700, carfentanil) emerged on the drug market. Recent research and several reports are pointing at the role of cryptomarkets in distributing novel psychoactive substances [3,4]. The importance of cryptomarkets has been further exacerbated by the spillover mental health and anxiety resulting from the ongoing COVID-19 pandemic: recent results from the Global Drug Survey suggest that the percentage of participants who have been purchasing drugs through cryptomarkets has tripled since 2014, reaching 15% of the 2020 respondents [5]. In this study, we assess social media data from active opioid users to understand the behaviors associated with opioid use and to identify what types of feelings are expressed. Substance use disorder (SUD) in social media posts is defined as a post that shows the risk of substance use, attitudes, and behavior related to substance use, as well as the corresponding social and environmental factors [6]. We used deep learning models to perform sentiment and emotion analysis of social media data with the drug entities derived from cryptomarkets. We implemented state-of-the-art sentiment and emotion models for social media data. Additionally, we performed topic analysis to extract frequently discussed opioid-related topics in social media. For the preliminary analysis, we examined temporal variations in topics that differentiate between posts at each drug level and topics over time across all years, followed by considering data per quarter for each year. We also analyzed how users' language in their posts varies temporally by topic. We also observed variations in emotions and sentiment that differentiate between posts containing expressions of SUD. For this task, we fine-tuned a pretrained transformer language model for emotions and sentiments, used it to automatically extract the emotions and sentiments for all historical posts related to a drug, and analyzed variations in sentiment and emotion over time. We further aim to achieve the identification of SUD on social media by examining the core research question of this study: can we differentiate between posts containing expressions of substance misuse or not with temporal activity, emotion, sentiment, and language features related to that drug? We built a knowledge-aware bidirectional sequential neural model that differentiates between posts where expressions of SUD are present versus those posts where it is absent.

Findings and Contributions

The major contributions and findings of this work are as follows:

- We compiled a high-quality, rare, challenging, and valuable dark web data set (eDark) by crawling four cryptomarkets, namely, Dream, Tochka, Agora, and Wall Street. The data set is available for release upon acceptance.
- We propose an end-to-end architecture (dark web to social media) for harnessing social media trends for opioid listings found on the cryptomarket. It involves crawling techniques, drug identification, data collection, processing from social media, and computational models to predict SUD considering the temporal variations in sentiment and emotional language among posts indicative of SUD. We also contribute to the knowledge- and historic posts–aware sequential neural model that can differentiate if SUD is present or absent for a drug based on these variations by factoring in the relative time difference between historical posts. We present that knowledge-, sentiment-, and emotion-aware models outperform other models of language feature–based approaches by performance measures, ablation study, and error analysis.
- To the best of our knowledge, our work is the first one to detect SUD in social media posts considering the above factors and as a reflection of the opioid listings extracted from the dark web. Resources created as a part of the study will be made available upon request to the corresponding author upon acceptance. The resources include emotion-, sentiment-, and SUD-labeled data sets with timestamps for each drug type, and the eDark data set.

Related Work

Dark Web Marketplaces

The dark web serves as a favorable and promising market for illegitimate goods ranging from drugs to weapons [7-9]. ElBahrawy et al [10] investigated the market dynamics of dark web markets based on a unique data set of Bitcoin transactions. They have also analyzed how the market ecology restructures itself once it closes. As traditional web scraping tools have failed to remove the veil of the vendors of dark marketplaces, Hayes et al [11] proposed an automated framework to overcome this barrier. The suggested framework was further evaluated by gathering information from 3000 sellers on a dark marketplace. Harviainen et al [12] presented an analysis of the pattern that the buyers and the sellers expose themselves on Sipulitori (a Finnish dark web drug trading market). Hassio et al [13]

extended research on Sipulitori by exploring it from the viewpoint of understanding the needs behind the messages posted by users and the physiological and cognitive factors that come into play. Researchers examined the underground marketplaces Agora and Dream Market to examine fluctuations in the availability of fentanyl, fentanyl analogs, and other illegal opioids in connection to overdose fatalities [14]. Orsolini et al [15] provided intuition behind dark web drug marketplaces through the perspective of psychiatrists so that they can be equipped with adequate information for providing countermeasures to increasing addictions to drugs available through these marketplaces. The prior work on analyzing dark web marketplaces suggests that such data could detect trends in the real world. Next, we discuss how time series analysis on social media helps to quantify such trends.

Time Series Analysis on Social Media

Earlier research has demonstrated the use of time series analysis on social media data, such as for comprehending changes in the sentiment of the public's perceptions, which can be beneficial to the government and commercial organizations [16], and understanding the sentiments of users of addictive smartphone apps such as PUBG and TikTok [17]. Time series analysis has also been used in research on mental health [18], such as variations in individuals' mental health throughout the COVID-19 lockdown phase [19]. A study conducted among adolescents in Nigeria examined how peer pressure and substance use affect the mental health of in-school adolescents, finding significant associations between these factors [20]. It highlights the need for interventions addressing peer pressure and substance use to promote positive mental health outcomes among adolescents. Over time, topic analysis and sentiment analysis have been used to deepen the understanding of web-based retail customer behavior from tweets [21]. Researchers have used time series analysis to analyze bursts of activity in social networks, and for prediction, they have used a long short-term memory (LSTM) network–based model [22]. A sparse additive generative model, a topic analysis tool, was used to assess the temporal linguistic changes in tweets with and without evidence of self-harm. Furthermore, they explored temporal linguistic features of tweets with and without suicidal intent signs [23]. A transformer-based model was also proposed for suicidal ideation detection in social media that takes into consideration the temporal context [24].

Substance Use Analysis on Social Media

Several researchers have explored social media analysis for different investigations of drug use. These works have analyzed the content, sentiment, and emotion for drug-related data collected from social media platforms like Twitter and Instagram. Lossio-Ventura and Bian [25] worked on a large amount of opioid-related data collected from Twitter to gain an overall understanding of drug-related discussions on the platform, behavior related to drug consumption, drugs co-used, and street terms for various drugs. This study restated that Twitter had a huge corpus of data and could provide insights into its correlation with pain management and alcohol consumption. A similar study by Cherian et al [26] was conducted on Instagram data on the misuse of codeine. The temporal data collected related to codeine misuse showed its interconnection with alcohol and soda consumption. The influence of social media in propagating this imagery increases the risk of normalizing drug use to extremes. Kim et al [27] further explored how big data can be used to understand drug use and addiction better. Social media is a large platform for monitoring prescription drug use and addiction using linguistic and behavioral cues. The work done by Lokala et al [14] investigates the relation between the availability of fentanyl-related drugs on cryptomarkets on the dark web and overdoses of fentanyl. Time-lagged correlation analysis was done between fentanyl-related drugs from the cryptomarket and overdoses of fentanyl in this first-of-its-kind study for epidemiological surveillance. Sarker et al [28] investigated various opioid-related subreddits to better understand the differences in conversations concerning prescription/illegal opioids and access to SUD treatment during the pre–COVID-19 and COVID-19 periods. They also noticed a rise in opioid withdrawal discussions during COVID-19. Posts from various subreddits related to opioids (both medical and illicit) were collected to identify the increase in the use of stimulants among opioid users and individuals with opioid use disorder [29]. This corresponds to the increasing number of casualties because of opioid and stimulant overdoses. Desrosiers et al [30] reported the perseverance of negative sentiments in the conversations of individuals with increased drug use severity. Liu et al [31] outlined the presence of positive emotion in the Facebook posts of individuals who underwent SUD treatment for a longer period than those who stopped their therapy. A study was also conducted by Singh and Wu [32] to probe the sentiment patterns of tweets related to SUD before and during the COVID-19 pandemic. Cameron et al [33] followed the development of a semantic web platform called Prescription Drug Abuse Online Surveillance and Epidemiology (PREDOSE) for harvesting data related to prescription drug use from social media platforms. Supporting several types of content analysis, PREDOSE provides easy access to data for drug use research. Fan et al [34] illustrated a new framework called AutoDOA to detect drug addiction behavior from Twitter. This will aid in understanding patterns of drug use and addiction. Eshleman et al [35] discussed how social media can be leveraged for drug recovery. Using linguistic patterns and machine learning algorithms, identifying groups of people more likely to participate in the drug recovery process would be an important step in managing the drug addiction epidemic. Our work aims to build an end-to-end system where we can see the reflection of the dark web on social media in terms of trends, sentiment, emotion, and substance use context, which is necessary for timely public health interventions.

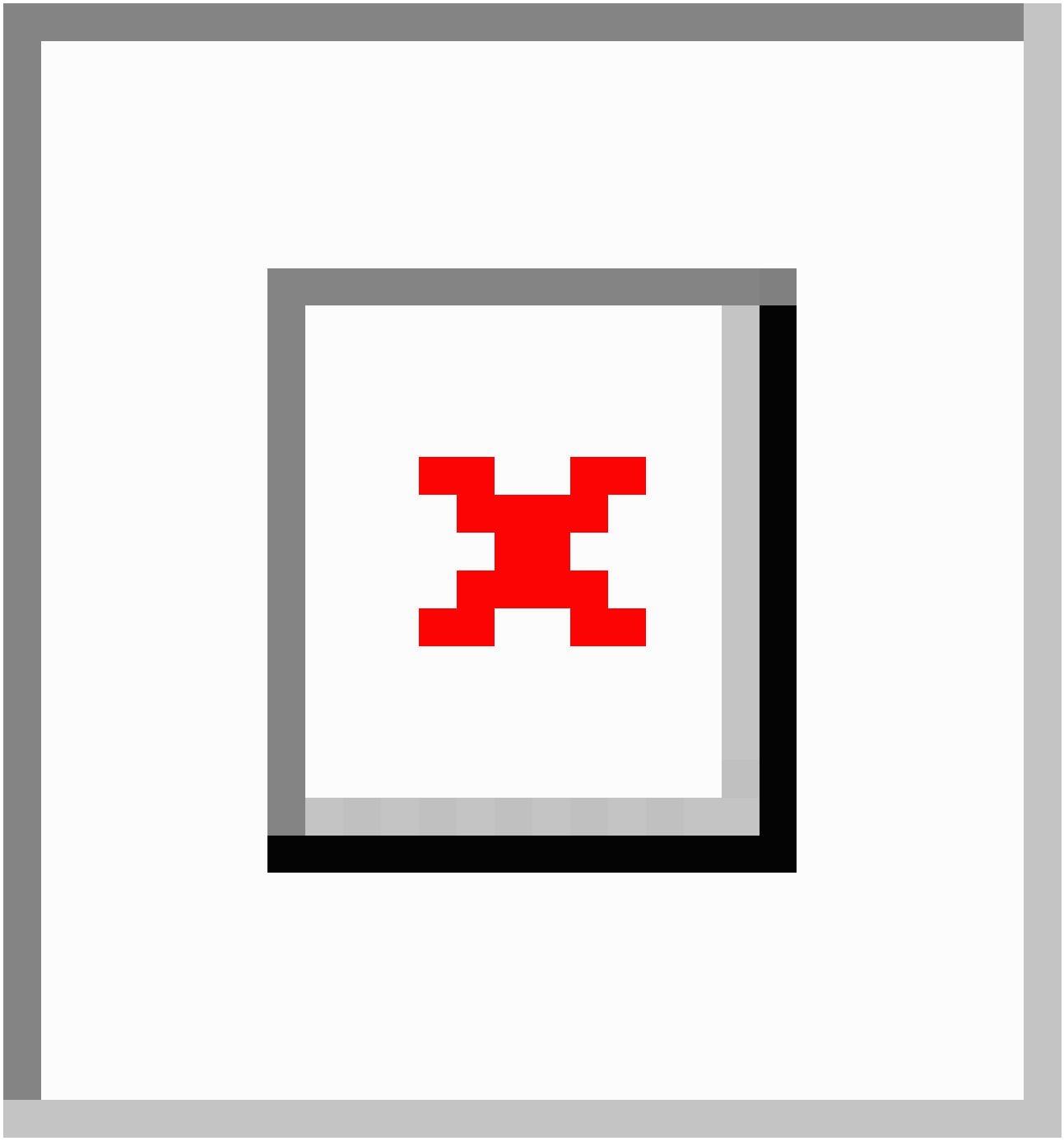
Data Collection

This section presents the modules-crawling techniques, drug identification, and data collection proposed in the dark web to social media architecture as shown in Figure 1.

Concerning dark web data, four cryptomarkets, Agora, Dream Market, Tochka, and Wall Street, were periodically crawled between June 2014 and January 2020. Over 82,000 opioid-related listings were collected to extract posts about fentanyl, fentanyl analogs, and other nonpharmaceutical

synthetic opioids in the cryptomarket. The data sources include four different cryptomarkets. Further, we discuss the eDark data set summary and a description of cryptomarkets in this section.

Figure 1. Proposed dark web to social media architecture for harnessing social media trends for listings found on the cryptomarket eDark data collection. DAO: drug abuse ontology; NER: named-entity recognition.



Dark Web Data (eDark) Collection and Summary

Dream Market

The marketplace was established in late 2013. Dream Market, after AlphaBay, was the largest dark web market in the world before 2017. Nevertheless, Dream Market quickly overtook AlphaBay as the largest dark web market in the world once AlphaBay went down in 2017 [36]. Between November 2014 and April 2019, there were 261 withdrawals from the market in total. During this time, the market saw transactions worth over US $197,000 [37].

Tochka

The marketplace started operating in 2015. It is a fairly modest market that mostly operates in North America and Europe. More than 3621 listings, including pharmaceuticals, malware, and other products, are sold on the website. The market changed its name to the Point Market and is currently open [38]. Between November 2014 and April 2019, there were a total of 2990 withdrawals from the market.

Wall Street

The marketplace features a site for the sale of illegal substances, weapons, hacking tools, and stolen log-in information. However, the exit scam has been hurting the market since April 2019 [39]. The administrators allegedly stole between US $30 million worth of XMR and Bitcoins from vendor accounts by switching the site into maintenance mode and transferring the clients' funds [40]. In May 2019, the market was later shut down. Before being taken over in May 2019 by the German Federal Criminal Police, Wall Street was the second-largest dark web market in the world. In total, 7755 withdrawals were made from the market between November 2014 and April 2019. During this time, there was almost US $18,000 worth of transactions on the market [37].

Agora

The marketplace was a dark web market in operation from September 2013 to August 2015 that sold illegal narcotics and controlled substances, drugs, counterfeit and fraud-related goods, services, and other illegal contraband. The data for Agora from June 2014 to September 2015 was obtained from the Grams data set [9]. Agora was chosen because it was one of the largest cryptomarkets that emerged after the Federal Bureau of Investigation shut down of Silk Road [41].

Data Summary

The summary of the dark web data set is shown in Table 1.

A sample product page of Dream Market is shown in Figure 2. The Scrapy framework was used to create the unique web crawler for each market, circumventing security protections built into these markets. To get over security safeguards, the web crawler uses specialized Scrapy downloader middleware. By creating a Linux virtual machine on Amazon Web Services running the Tor daemon and Privoxy, the custom crawler was able to reach the deep web. The outputs of the crawler are unaltered HTML files used for drug advertising. The university's information security office evaluated and approved the data extraction, storage, and access processes, which all adhered to stringent security standards. The information that was extracted from the data included the following: the product name provided by the vendor, the vendor screen pseudonym, the number of sales made by the vendor and their level of trust, the drug names, drug category, the information the vendor provided about the product, the unit, the quantity in stock, the price (in Bitcoin and US dollars), the price per volume, the country/region of origin, the destination country/region, and the security precautions for transactions. We further used a custom-built named-entity recognition (NER) algorithm to extract substance names, product weight, price of the product, shipment information, availability, and administration route as shown in Table 2. The NER algorithm consists of three key components: (1) the Natural Language Toolkit is used to curate and process text portions from crawled data; (2) the drug abuse ontology (DAO) that serves as a conceptual framework for interconnecting groups of drug-focused lexicons to produce a list of items to be identified; and (3) regular expressions, which are a sequence of symbols and characters that create a pattern that can be searched in text or a sentence constructed using the DAO-selected entities to extract things of interest.

Table . eDark summary.

Data	Dream Market, n	Tochka, n	Wall Street, n	Agora, n
Vendors	3456	765	876	910
Substances	2862	679	765	821
Locations	436	62	37	214
Dollar worth (US $)[a]	197,000	5072	18,000	220,000
Withdrawals[a]	262	2990	7755	844

[a]US dollar values and withdrawals are approximated to the nearest value.

Figure 2. Data source of eDark: a sample product listing page from the Dream Market cryptomarket.

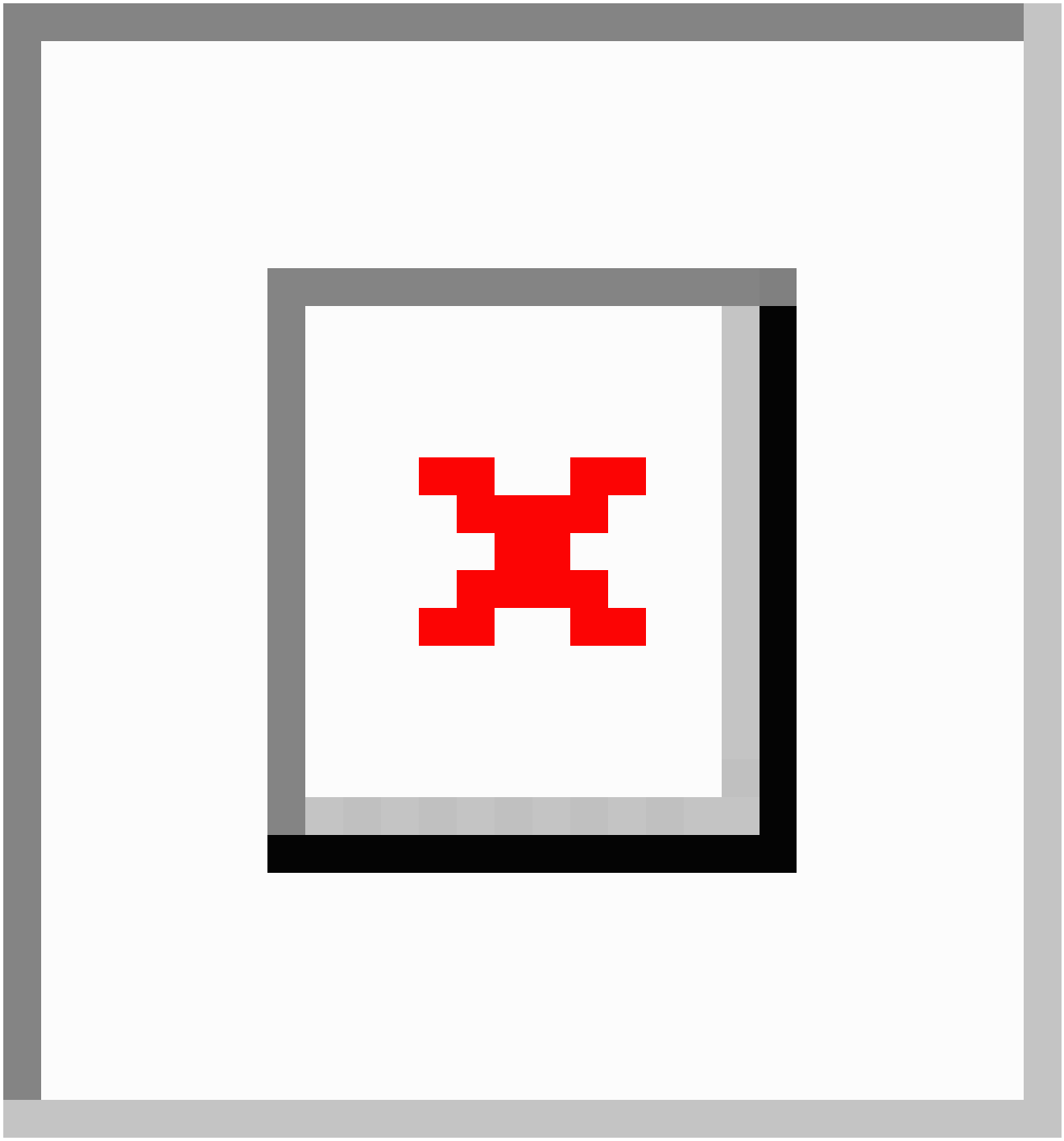


Table . Sample of property types in eDark identified from cryptomarket product listing.

Property name	Cryptomarket listing information
Has product name	50 Gr ***** Heroin AAA +With Spots Free Shipping
Is substance	Heroin
Has class	Opiate
Has dosage	1.5 gram
Has quantity	50 gram
Has vendor	BulkBrigade
Has price	BTC 0.0444
Ships to	Worldwide
Ships from	Germany

Named-Entity Recognition

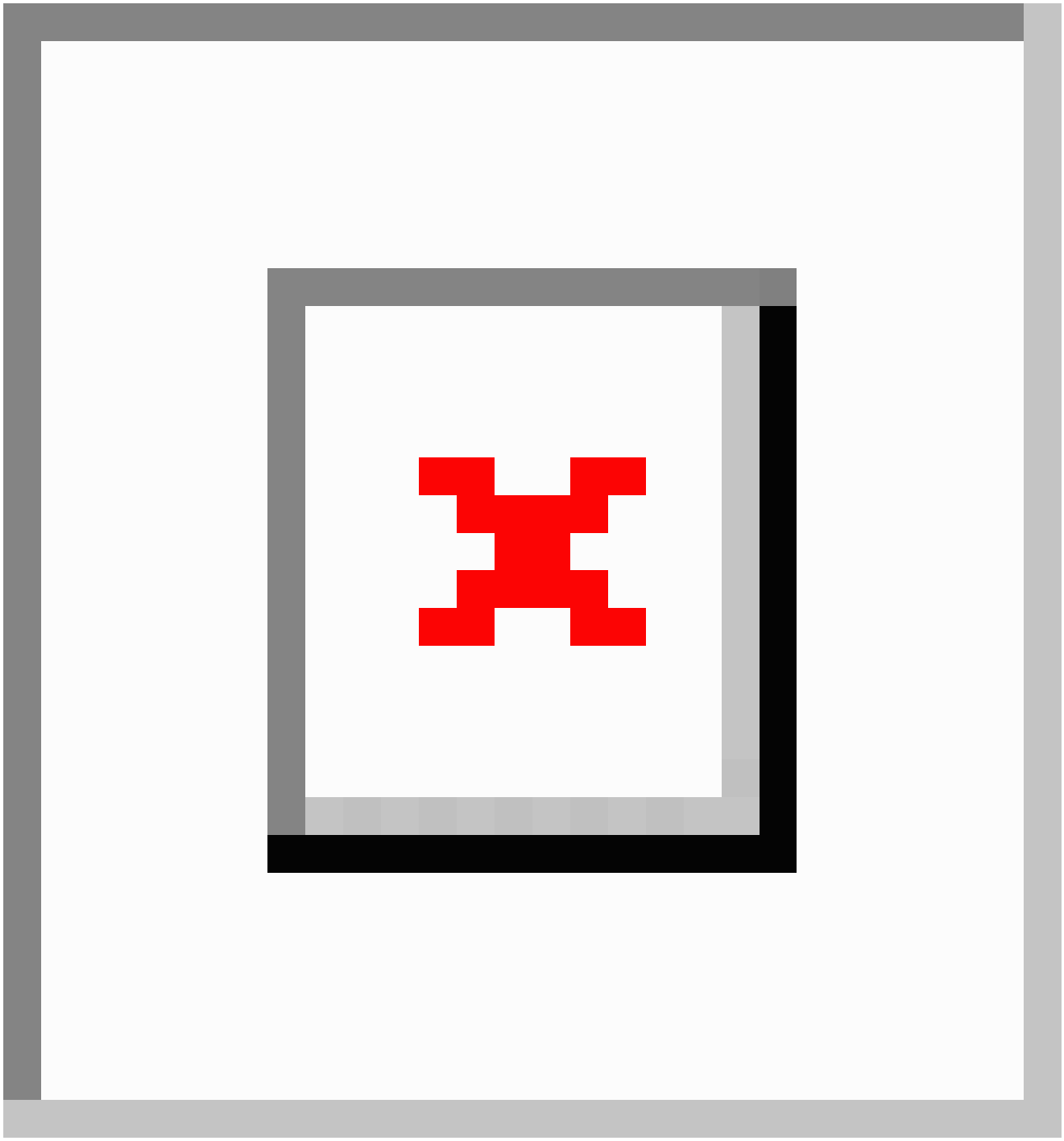
Extracted data included features like product name, vendor screen name (vendor name), drug category, product description, price (Bitcoin or US dollar), country/region of origin and destination, how to administer the drug, shipping information, and others. We used a pretrained NER deep learning bidirectional LSTM–convolutional neural network (CNN) approach [42] on cryptomarket data to identify drug entities that use a hybrid bidirectional LSTM and CNN architecture, eliminating the need for most feature engineering. The entities are then matched to a superclass using DAO [43] that acts as a domain-specific resource. DAO is a domain-specific knowledge source containing drug- and health-related classes, properties, relationships, and instances. Apart from medical terms, it includes concepts of mental health disorders and symptoms aligned with the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) scale. DAO is a domain-specific conceptual framework for interconnecting sets (named "classes") of drug-focused lexicons. One of the key benefits of using an ontology-enhanced semantic approach is the ability to identify all variants of a concept in data (eg, generic names, slang terms, scientific names). The DAO contains names of psychoactive substances (eg, heroin, fentanyl), including synthetic substances (eg, U-47700, MT45), brand and generic names of pharmaceutical drugs (eg, Duragesic, fentanyl transdermal system), and slang terms (eg, roxy, fent). It also contains information regarding the route of administration (eg, oral, intravenous), unit of dosage (eg, gr, gram, pint, tablets), physiological effects (eg, dysphoria, vomiting), and substance form (eg, powder, liquid, hydrogen chloride). Initially, it was used to determine user knowledge, attitudes, and behaviors related to the nonmedical use of buprenorphine and other illicit opioids through analysis of web forum data. Later, this ontology evolved to understand trends of drug use in the context of changing legalization policies in the United States. This also proved effective in capturing gleaning trends in the availability of novel synthetic opioids through the analysis of cryptomarket data. DAO is defined as using a common ontology methodology known as 101 ontology development. The 101 technique entails the following steps: (1) establishing the ontology's domain and scope, (2) reusing prior knowledge, (3) enumerating key terms in the ontology, (4) defining classes and their properties, and (5) producing instances of the classes. A collection of techniques and best practices accepted by the Semantic Web community and the artificial intelligence community that perform natural language processing were used to assess the ontology's quality. Protege is the most used tool for creating ontologies [44], so the metrics list the numbers for its structures and representation. The DAO metrics were evaluated as shown in Table 3.

The Web Ontology Language representation of DAO is presented in Figure 3. In this study, we leveraged DAO to identify 90 drug entities, which we then broadly classified into 8 categories by mapping each entity to a super drug class in DAO. The 8 broad categories considered were heroin, synthetic heroin, pharmaceutical fentanyl, nonpharmaceutical fentanyl, fentanyl, oxycodone, kratom, and opium (chosen as per data availablity in each category on social media). The categorization of the 5 types of opioid listings containing specific types and subclasses identified using DAO is shown in Textbox 1.

Table . Drug abuse ontology metrics.

Ontology metric	Value, n	Description
Axiom	4876	Number of combined logical and nonlogical axioms
Logical axiom count	3478	Number of logical axioms
Declaration axiom count	1185	Number of declaration axioms
Classes	316	Number of distinct classes
Objects	12	Number of object properties
Data property	13	Number of data properties
Individual count	845	Number of individual entities

Figure 3. The Web Ontology Language representation of drug abuse ontology.



Textbox 1. Opioid listings categories, subclasses, and specific types identified using drug abuse ontology.

Pharmaceutical fentanyl

Duragesic, Sublimaze, fentanyl transdermal system

Nonpharmaceutical fentanyl

Oxycodone pills with fentanyl

Fentanyl analogs

Acetylfentanyl, acrylfentanyl, alfentanyl, benzylfentanyl, betahydroxyfentanyl, betamethylfentanyl, butryfentanyl, carfentanil, crotonylfentanyl, etorphine, etorphinecartanil, fluorofentanyl, isobutyrfentanyl, lofentanyl, methoxyacetylfentanyl, methylfentanyl, etc

Novel synthetic opioids

U-50488, U-47700, U-49900, U-48800, MT-45, AH-7921, W-18, MPF-47700, etc

Pharmaceutical opioids

Buprenorphine, codeine, hydrocodone, hydromorphone, loperamide, methadone, morphine, naloxone, oxycodone, oxymorphone, tramadol

Identifying Substance Use Discussions on Social Media

We crawled the data using a carefully curated lexicon extracted from DAO consisting of around 120 terms (slang names, brand names, drug names, street names, marketing names, commonly used names, and abbreviations) of those 8 drug categories.

Using the compiled list, we collected 290,458 opioid-related posts from 6 subreddits using custom-built crawlers, which we call the Substance Use Disorder Corpus (SUDS). The 6 subreddits chosen for data collection were r/drug nerds, r/research chemicals, r/opiates, r/heroin, r/suboxone, and r/opiates recovery. The subreddit corpus is spread over different drug categories such as heroin (n=136,745), kratom (n=77,443), fentanyl (n=36,166), oxycodone (n=25,890), opium (n=9675), nonpharmaceutical fentanyl (n=2798), pharmaceutical fentanyl (n=876), and synthetic heroin (n=865). To build the social media emotion analysis model, additionally, we collected 151,563 posts from Twitter using the Twitter application programming interface with the same lexicon we used for the subreddit crawler. We applied term frequency–inverse document frequency (TF-IDF) over unigrams, bigrams, and trigrams to identify topics in each subreddit as shown in Textbox 2. We also conducted the topic analysis using the BERTopic [45] model for all drugs over time from 2015 to 2020, as shown in Figure 4.

Textbox 2. Sample of topics identified from the Substance Use Disorder Corpus data set obtained from 6 different subreddits.

Opiates recovery

Cold turkey withdrawal, cravings, anxiety, rehab, depression, sobriety, loperamide, benzo, Subutex, quitting, Vivitrol, Imodium, naltrexone

Opiates

Codeine, hydrocodone, oxymorphone, Dilaudid, hydromorphone, Opana, OxyContin, acetaminophen, gabapentin, benzos, Roxicodone

Suboxone

Buprenorphine, Subutex, agonist, clonidine, tramadol, hydrocodone, Dilaudid, Vicodin, Sublocade, Percocet, phenibut, Klonopin, Valium

Heroin

Dope, opium, opiates, crack, diacetylmorphine, China white, codeine, acetaminophen

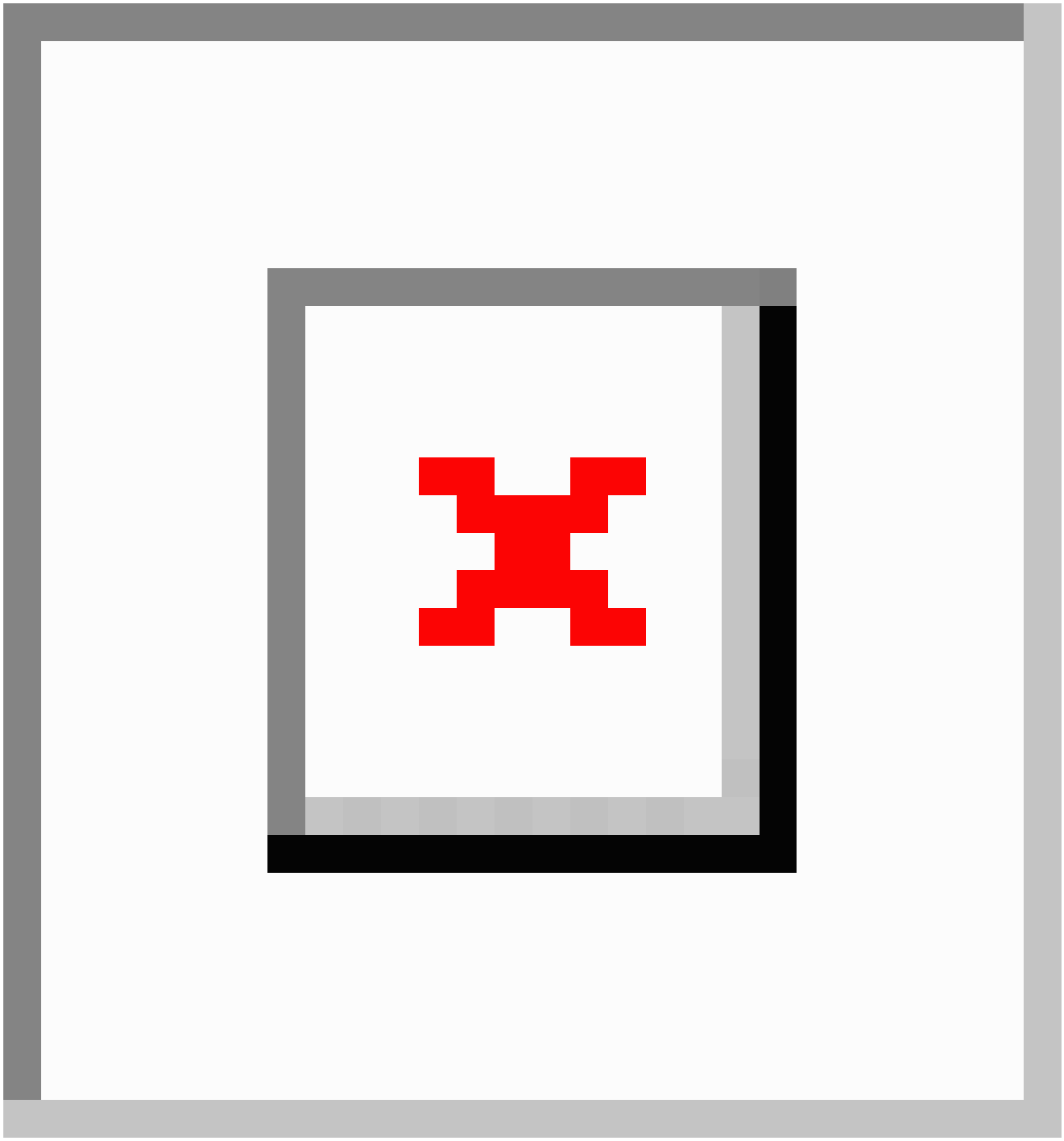
Drug nerds

Methadone, alkaloids, mitragynine, benzos, poppy, buprenorphine, antagonist, gabapentin, naloxone, amphetamine, hydrocodone

Research chemicals

Benzos, psychoactive, psychedelic, kratom, pyrovalerone, quaalude, oxycodone, morphine, Xanax, tramadol, cocaine, methadone, ketamine, gabapentin, amphetamine, hydromorphone

Figure 4. Dark web to social media topic modeling: module topics over time for all 8 drug categories from 2015 to 2020.



Methods

Overview

In this section, we build upon the previous data collected to create Bidirectional Encoder Representations From Transformers (BERT)–based sentiment, emotion, and SUD models. We leveraged those models to predict if a post could be classified as SUD present (SUDP) or SUD absent (SUDA) while considering the history of the post. We applied stratified random sampling [46] to identify a sample population that best represents all the features of interest and ensures that every data subgroup is represented, thus avoiding potential bias in the several data sets we collected for this study.

Sentiment Analysis and Sentiment BERT Model

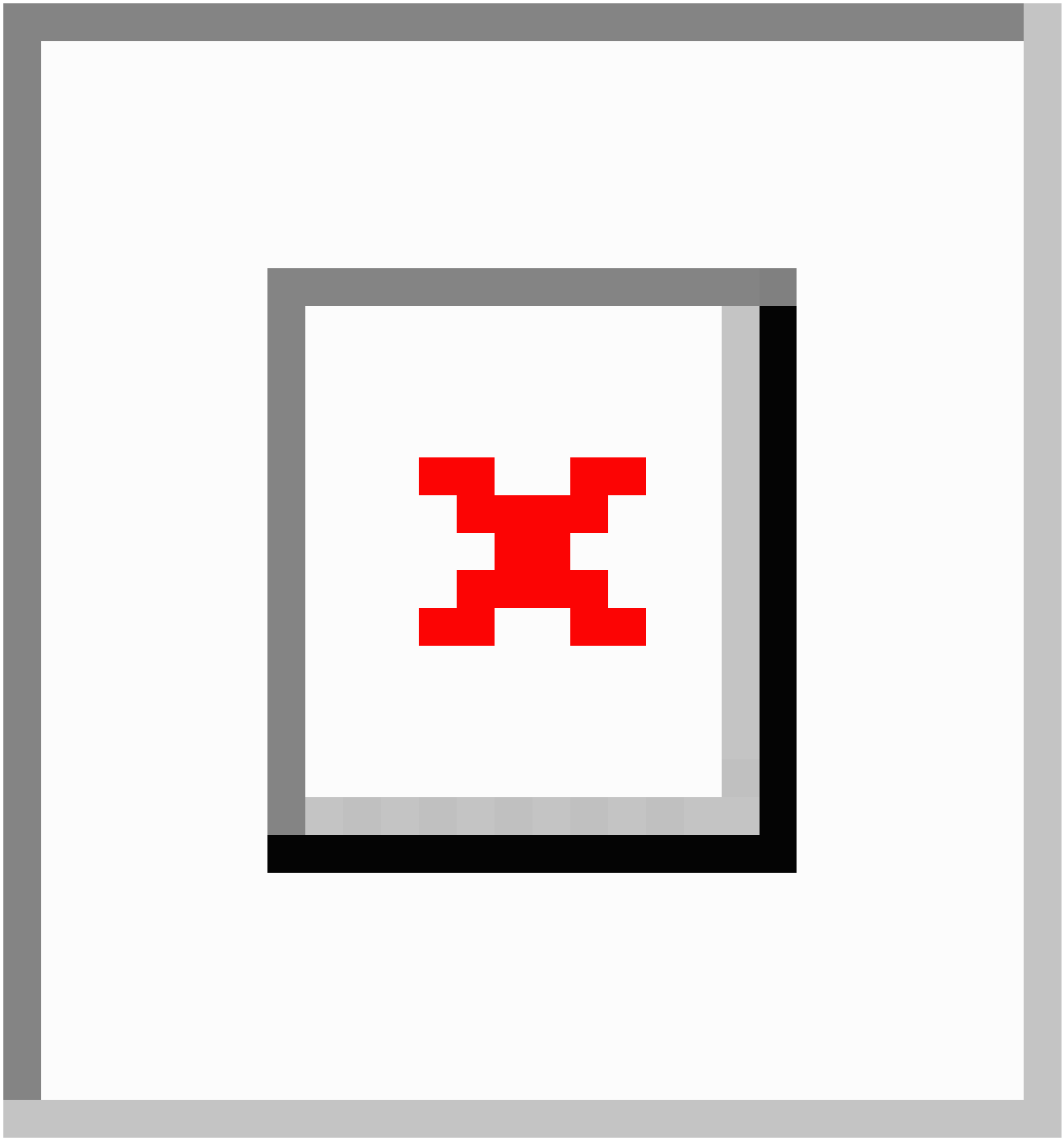
We classified subreddit posts as positive, negative, and neutral for the sentiment analysis. We implemented Valence Aware Dictionary for Sentiment Reasoning (VADER) [47] to generate sentiment for each subreddit post in the SUDS to consider both the polarity and intensity of each sentiment. VADER uses a lexicon of words with human-annotated sentiment polarity scores like SentiWordNet, AFINN, and the National Research Council Canada Word-Emotion Association Lexicon. We chose VADER as it is a rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and it uses a combination of sentiment-related words, emoticons, and syntax to produce a sentiment score for a given text. Following the individual scoring of each word, the ultimate

sentiment is determined by performing a pooling procedure, such as averaging all the sentiments. This data set is split into train, dev, and test sets (75:5:20). The generated training set is used to train state-of-the-art deep learning algorithms like CNN, LSTM, and BERT. The highest F_1-score achieved was 82.36 with the BERT model. We trained the sentiment BERT model on this training data for later use. We report the statistics of sentiment labels for subreddit posts obtained from sampling 800 random data points from each drug category in Table 4. The comparison of the drug categories of pharmaceutical opioids and heroin by the top three sentiments—positive, negative, and neutral—for the period between 2015 and 2020 is presented in Figure 5, which shows the temporal variation in sentiment for each drug.

Table . Dark web to social media sentiment and emotion analysis module-sentiment stats (number of posts) after sampling 800 random points for each drug category identified from 6 subreddits and the top emotions identified for each drug from Twitter.

Drug	Positive, n	Negative, n	Neutral, n	Top 3 emotions
Opium	481	218	101	Sadness, love, joy
Oxycodone	460	245	95	Sadness, fear, thankfulness
Kratom	459	231	110	Love, sadness, fear
Fentanyl	467	274	59	Sadness, love, fear/thankfulness
Heroin	455	255	90	Sadness, joy, thankfulness
Synthetic heroin	500	240	60	Sadness, fear, thankfulness
Pharmaceutical fentanyl	570	197	33	Sadness, love, joy/thankfulness
Nonpharmaceutical fentanyl	502	264	34	Sadness, love, thankfulness

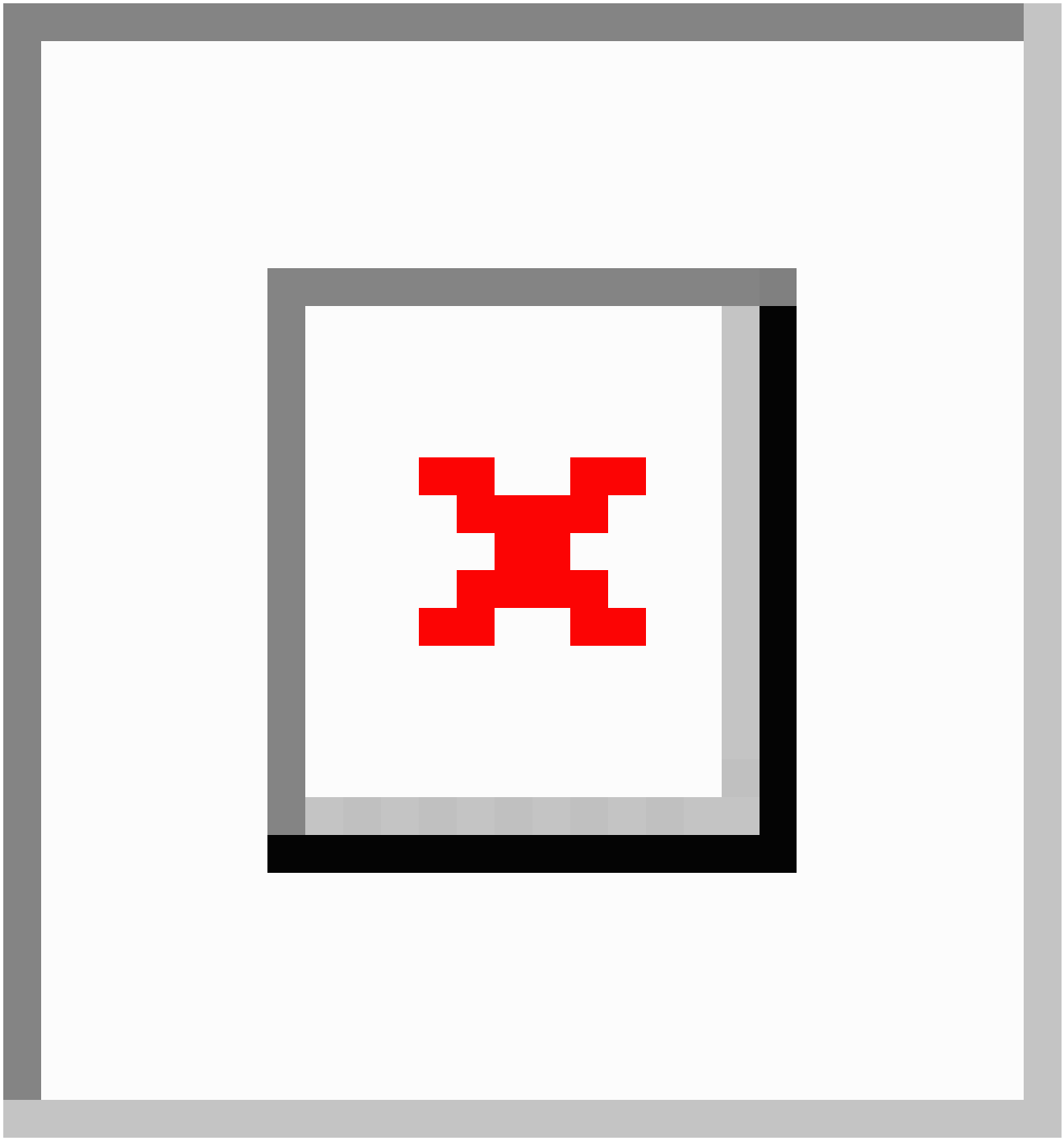
Figure 5. Dark web to social media sentiment analysis module comparison of the drug categories pharmaceutical opioids and heroin by sentiments.



Emotion Analysis and Emotion BERT Model

We did not choose to work on subreddit data for emotion analysis as we did not have self-tagged emotions in posts on the subreddits. Therefore, we chose to crawl Twitter for the emotion analysis, where emotions are presented as hashtags. We limited our crawl to 7 emotions, as stated by Wang et al [48]. The tweets are assigned a class label corresponding to the emotion hashtag they are associated with. Furthermore, we removed any URLs or usernames that could potentially contain sensitive information. For generating emotion labels for drug-related tweets, we implemented an inductive transfer learning approach with BERT [49]. For this task, we extracted 61,000 posts as labeled training data by crawling tweets with each emotion hashtag: joy, sadness, anger, love, fear, thankfulness, and surprise. We split this data set into train, dev, and test sets (75:5:20). We trained Emotion BERT, which is a BERT-based model for 10 epochs using a learning rate of 1e-5, on a batch size of 32 of the labeled data and on Emonet, a corpus of around 790,000 tweets [50], to generate the emotion labels for subreddit posts in the SUDS. We report the statistics of emotion labels for subreddit posts obtained from sampling 800 random data points from each drug category reported in Table 4. The comparison for all drugs across the 7 emotions: joy, sadness, anger, love, fear, thankfulness, and surprise is shown in Figure 6.

Figure 6. Dark web to social media emotion analysis module comparison of the 8 major drug categories across 7 emotions: joy, sadness, anger, love, fear, thankfulness, and surprise.



Substance Use Disorder Data Set

We focused on building and interpreting a predictive model based on these exploratory results to identify posts where SUD is present or absent. We formulated this problem as a binary classification task to predict a label for a post at a particular time. Each post was associated with a drug name, historical posts, time, emotion, and sentiment. We then prepared our training data set to generate SUDP and SUDA labels for the SUDS. We made use of high-quality addiction-labeled data from Lokala et al's [51] work on social media data for exploring the association between drug and mental health symptoms. Lokala et al [51] created a labor-intensive high-quality corpus of 9888 tweets manually annotated by domain experts and substance use epidemiologists with experience in interventions, treatment, and addiction research. We trained a transfer learning BERT model for 10 epochs using a learning rate of 1e-5 and batch size of 32 on this labeled data to generate the SUDP and SUDA labels for posts in the SUDS. We also examined manual interannotator agreement (κ=0.74) among three domain experts for SUDP and SUDA labels of 300 posts to validate the annotations. The manual annotations were evaluated in the same way as automated labels, and our macro–F_1-score measure against the ground truth was 0.71. The results for transfer learning using BERT are reported in Table 5.

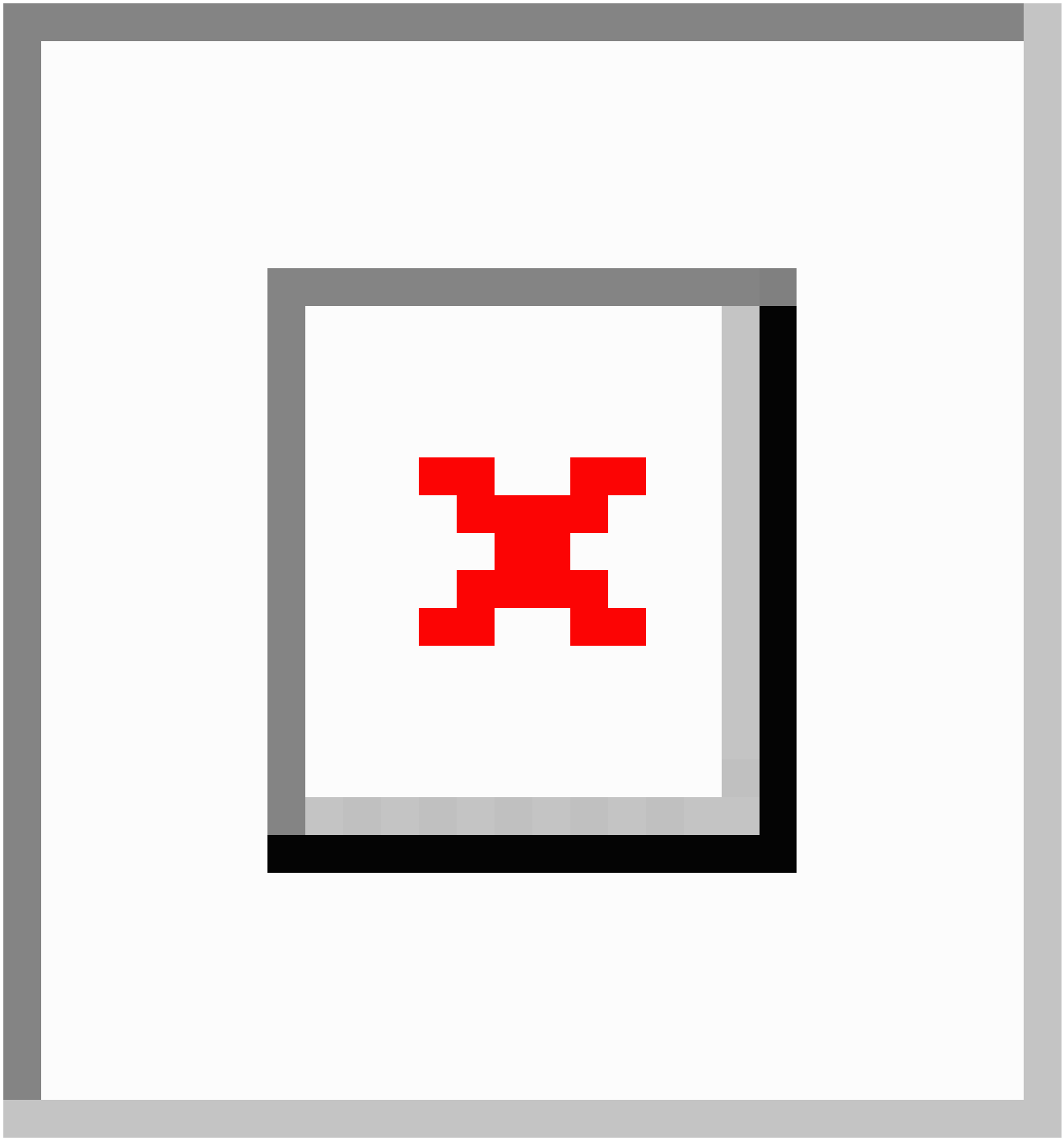
Table . Validation results for emotion Bidirectional Encoder Representations From Transformers (BERT) and substance use disorder (SUD) BERT models through the transfer learning approach. The trained model was then used to obtain the emotion labels SUD present and SUD absent for the posts in the Substance Use Disorder Corpus.

Transfer learning BERT model	Precision	Recall	F_1-score
Emotion BERT	80.12	82.29	81.19
SUD BERT	81.28	83.65	82.44

Temporal Predictive Model of Posts to Detect SUDP or SUDA

We built our domain-specific sentiment BERT model to serve as a sentiment feature extractor over historical tweets and a domain-specific emotion BERT model as an emotion feature extractor for historical tweets. We built fine-tuned BERT models as they can capture a better sense of sentiment, emotions, social media jargon, and slang terms [52]. We contributed to the knowledge-aware time series analysis computation model to predict SUDP and SUDA for a post as shown in Figure 7. We presented the SUD detection as a binary classification model with SUDP and SUDA as labels. We focused on building and interpreting a predictive model based on these exploratory results to identify posts where SUD was present or absent. Each post was associated with the drug, historical posts, time, emotion, and sentiment in the SUDS.

Figure 7. Knowledge-aware time series analysis computation model: a dark web to social media computational module. BERT: Bidirectional Encoder Representation From Transformers; SUD: substance use disorder.



For a post, the concepts, slang terms, and synonyms related to drug entities are masked using the DAO, which forms the knowledge component of the model. We used BERT to encode the language representation, as BERT can produce more thorough representations of linguistic elements in social media data [52], and we averaged the vector outputs for all tokens in each post at the final layer. To extract emotional language features from posts, we used our emotion BERT model that takes the historical posts and obtains the 768-dimensional emotion vector of each historical post. To extract sentiment language features from posts, we used our sentiment BERT model that takes the historical posts and obtains the 768-dimensional sentiment vector of each historical post. We then have the encoding representing the sentiment and emotion spectrum. Sequential models such as recurrent neural network and LSTM models are apt ways to encode representations that learn from a sequence of a user's historical tweets due to the sequential nature of a social media post history. We then passed the historic posts through a bidirectional LSTM + attention layer concatenated with the post to be assessed. We then fed the extracted features from the attention layer to a dense layer with the rectified linear unit to get the prediction vector. Finally, we used the softmax function to output the probability that the post is labeled as SUDP or SUDA.

For experiments, we split the data set into a 75:5:20 ratio for the train, development, and test sets, respectively. We fine-tuned

the hyperparameters using the development set. Each model was trained for 10 epochs with a learning rate of 1e-5 and a batch size of 64. We used cross-entropy loss and Adam [53] for the optimization. For regularization, we used dropout [54] with a probability of 0.2. We got the best performance with the bidirectional LSTM model with the attention layer as it captured context over a longer span considering the bidirectional context of a word. For all models, we report the recall, precision, and F_1-score. We interpreted the higher performance gains of our model in the Results section through an ablation study.

Performance Comparisons

We compared the performance of these state-of-the-art methods through replications of the architectures and representations presented in prior works on similar tasks on social media.

- Logistic regression [55]: We implemented a logistic regression classifier that uses the part of speech (POS) and TF-IDF as language feature representations.
- Random forest [56]: We implemented a random forest model with features like Linguistic Inquiry and Word Count (LIWC), POS, and TF-IDF.
- History-aware recurrent neural network (H-RNN) [57]: We deployed H-RNN that encodes input using fine-tuned fast text embeddings. Historical posts are passed sequentially through the model and concatenated with the post to be assessed. The sigmoid activation was selected for the hidden LSTM layer, which is fully connected to both the input and output layers.
- History-aware LSTM (H-LSTM) [58]: We replicated H-LSTM that used BERT embeddings for encoding historic posts given to an attention-based LSTM layer, allowing the model to choose whether to focus more or less on each post to reflect user representation, which was finally be fed to a fully connected layer with a sigmoid activation function to get the prediction.

Ethical Considerations

We applied our model to study how historic emotion and sentiment of a drug impacts social media conversation dynamics related to substance use. An important aspect that we need to consider while working with addiction-related issues is to respect the users' privacy and adhere to good ethical practices adopted by previous research [59,60]. Therefore, similar to Matthews et al [61], we censored sensitive information such as usernames, personal information, platform identifiers, and URLs that might be directly linked to the user's identity from the collected posts. All examples used in this paper are anonymized and deidentified for user privacy [62]. We also adopted the proposed guidelines for the treatment of names and online pseudonyms in posts gathered from social media [63]. In this work, we study substance use in subreddit groups in the form of textual interactions. The expressed addiction intent may differ from the intent perceived or experienced by the person. However, obtaining perceived intent from social media is challenging and involves ethical risks. Before behavioral health intervention apps based on social media data can be used in real-world settings, difficulties with potential biases and user privacy must first be resolved, along with establishing suitable regulations and boundaries in this domain. We can adopt the approach used in this work to develop the data set with more human supervision, and we acknowledge that the data may be prone to demographic-, annotator-, and platform-specific biases [64,65]. We also acknowledge that this study does not make any clinical diagnosis or treatment suggestions directly. However, its findings can still be highly valuable for informing clinical practice and interventions for SUD in the identification of risk factors. Even if the study does not prescribe specific treatments, its findings can inform the development of novel intervention strategies or the optimization of existing ones. For example, as our paper highlights certain sentiment and emotion patterns associated with SUD, clinicians can incorporate this knowledge into their approaches. This research also provides further understanding of the factors contributing to the onset of SUD, which can help in designing prevention programs aimed toward at-risk populations, such as individuals with mental health disorders. Research findings can inform policy makers about the effectiveness of current strategies and the need for adjustments in regulations, health care policies, or resource allocation for addressing SUD at a broader level.

This study was submitted for review by the institutional review board at the Wright State University and has been determined to meet federal exemption criteria: 45 CFR 46.101(b)(4).

Results

Overview

Of all opioid listings in the eDark data set, 4.2% are related to novel synthetic opioids, and heroin was identified in 57.8% of all opioid-related listings. When comparing the average monthly ad volume for fentanyl, fentanyl analogs, and other nonpharmaceutical drugs, data indicates a rise in the availability of items containing fentanyl. The listings for pharmaceutical and nonpharmaceutical fentanyl and analogs made up 1.9% of all opioid-related listings, which is 48.6% of the unique synthetic opioid-related ads. The most frequent type of novel synthetic opioid, which was synthetic heroin, was offered for sale at an average of 1.6 kg at each time point of data collection during the study period. Furanylfentanyl was the fentanyl analog that was promoted the most with an average of 3.6 kg being offered for sale at each data point. Carfentanil, a highly strong fentanyl analog, was typically available for purchase for 489.6 grams on average. Newer synthetic opioids (eg, U-48800, U-4TDP) kept replacing the nonpharmaceutical synthetic opioids (eg, W-18, MT-45, AH-7921, U-47700) in the listings found on marketplaces.

From the exploratory analysis on the SUDS, kratom, heroin, fentanyl, morphine, cocaine, methadone, Suboxone, and oxycodone were the most commonly discussed drugs across 6 subreddits. In Textbox 2, considering the Research Chemicals subreddit, for example, it is interesting to find that more posts talk about pyrovalerone, a psychoactive drug with stimulant effects. Another term found is "Quaalude," a brand name for "Methaqualone," a sedative and hypnotic medication. The Research Chemicals subreddit mostly discusses psychoactive and psychedelic drugs, while DrugNerds discussed alkaloids [66]. Interestingly, DrugNerds talked about naloxone, which can treat opioid overdose. Dope is a slang term for heroin and

was identified in the Heroin subreddit. Several brand names of medications for anxiety, pain, seizures, insomnia, and sedatives are discussed in the Suboxone subreddit. Gabapentin is the typical seizure and pain medication discussed among most of the subreddits. The Opiates Recovery subreddit is more about withdrawal symptoms and mental health disorders, for example, "cold turkey." The term "cold turkey" used in the context of substance use is quitting a substance abruptly, which carries significant risks if the drug you are discontinuing is a benzodiazepine or opiate [67,68]. The results show that we can derive and analyze slang terms, brand names, novel drugs, mental health symptoms, and medications from social media. From the results in Table 4, the highest positive sentiment was found in pharmaceutical fentanyl, the highest negative sentiment in fentanyl, and the highest neutral opinion in kratom. The emotion "Love" was most detected in posts related to kratom as people use it for self-medication. Table 6 presents the medians of metrics for different embeddings and architectures obtained over 20 runs. The baseline models we compared our model to were logistic regression, random forest, H-RNN, and H-LSTM with varied language representations like POS and TF-IDF. We also extracted LIWC features from posts to pass through a predictive model instead of BERT encoding. Under identical circumstances, we empirically discovered that BERT outperformed LIWC considerably (P=.03). We presented model interpretability and significance as an ablation study.

Table . Median of metrics for different embeddings and architectures that were obtained over 20 runs.

Representation	Model	Macro–F_1-score	Precision	Recall
POS[a] + TF-IDF[b]	Logistic regression	52.72	47.67	50.31
Linguistic Inquiry and Word Count + POS + TF-IDF	Random forest	55.84	54.63	56.78
Fast text	History-aware recurrent neural network	74.47	69.48	76.17
BERT[c]	History-aware long short-term memory	76.85	70.56	77.61
Knowledge-aware BERT	Knowledge- and history-aware bidirectional long short-term memory + attention	*82.12*[d]	*78.34*	*83.58*

[a]POS: part of speech.

[b]TF-IDF: term frequency–inverse document frequency.

[c]BERT: Bidirectional Encoder Representations From Transformers.

[d]Italics denote that the result is significantly better than history-aware recurrent neural network (P=.04).

We used the Wilcoxon signed rank test [69] to compare the emotional expression in posts and comments between those with and without substance use to assess statistical significance. There was a significant correlation between the emotion displayed in posts labeled SUDP (P<.001) and the posts labeled as SUDA. We next conducted an ablation study where we removed one component from our model and assessed the performance to analyze the prime components in our methodology. We concatenated the substance use post encoding $e_i^{(S)}$ and emotion post encoding $e_i^{(E)}$, and used the resulting representation as the input to the linear layer to exclude the attention component from the model. We trained our encoders with raw data that was directly collected from social media to remove the entity-masking component. Additionally, we trained our model by merely training the classifier and excluding the post history from the model. In Table 7, we report the findings for the SUD prediction task for posts. Entity masking, which considerably improves the SUD identification task (+3.73 precision, +3.43 recall), is where we saw gains. The Wilcoxon signed rank test demonstrated that contextualized representation is very desirable for the SUD identification task in this study since it performs better than the model without entity masking (P=.04). Additionally, adding the history of the post significantly boosted the performance where we saw our highest increase (+4.62 precision). Attention also increased the model's precision by 1.85% and recall by 3.24%, meaning that every feature of the model affects how well it performs this task. Below, we discuss the examples of SUD and the result error analysis.

Table . Ablation study: median of metrics over 10 different runs.

Model	F_1-score, median (change from proposed model)	Precision, median (change from proposed model)	Recall, median (change from proposed model)
Proposed model (entity masking + attention + history of post)	*82.12* [a]	*78.34*	*83.58*
Minus attention	78.98 (3.14↓)	76.49 (1.85↓)	80.34 (3.24↓)
Minus entity masking	78.50 (3.62↓)	74.61 (3.73↓)	80.15 (3.43↓)
Minus history of post	77.58 (4.54↓)	73.72(4.62↓)	79.12 (4.46↓)

[a]Italics denote best performance.

Error Analysis

We analyzed the sources of errors and discuss the predictions made by our models in Table 8 in three scenarios.

1. Polydrug use with variable emotions: For post 1, examining the post where multiple drugs coexist along with emotion variability in the history associated with other drugs, for example, mixing depressants and stimulants or mixing medications with opioids, our model was not able to predict correctly, for example, when substance A might not often co-occur with substance B in the history.
2. Post-level ambiguity: For post 2, our model was able to predict SUD by examining the post even if it was too ambiguous to assess given that the user has a clear SUDP label in the past and was undergoing the healing process, with emotion intensity for the historical posts like increased sadness-related emotions.
3. Sarcasm detection: For post 3, even if it does not contain any clear SUDP/SUDA label, sarcasm identified in the post, with a history of ambiguous posts, presents difficulty in identifying SUD, which makes it an interesting natural language understanding problem and explains the task's complexity, revealing paths for future work.

Table . Examples showing major errors made by our proposed approach.

Post/comment	Error type	Actual	Predicted
1. "Imagine a combo. Just got stoned with 12 grams of kratom and 15 g of **** with 40 mg of ***** AMA lol, ended up projectile vomiting, went to sleep feeling fearful and woke up feeling pretty joyful but little shitty, so i had glass of ACV, all is well..."	Polydrug use	SUDP[a]	SUDA[b]
2. "The improved memory isn't dependent on the supplement. I can take the same dose of Nal****one to treat Opioid Use Disorder for 3 mo and not feel a thing. That tells me that it helped me heal, and there's not much more for it to do."	Post-level ambiguity	Ambiguous	SUDP
3. "why SPEAK the UNSPOKEN? Dangerously F***ing disaster. Mixing *** (a long-term damaging drug, with an ability to make other drugs dangerously stronger) with Xanax (addictive,overdoseable drug) and ***** (an opiate that can and likely will kill you) is a recipe for life after ***"	Sarcasm detection	Sarcastic	SUDA

[a]SUDP: substance use disorder present.

[b]SUDA: substance use disorder absent.

Discussion

Principal Findings

Crawling cryptomarkets poses a significant challenge when applied to data science and machine learning to study the opioid epidemic due to the restricted crawling process [1,43,70]. However, we make our proposed data set available to the research community for further analysis. To identify the best strategies to reduce opioid misuse, a better understanding of cryptomarket drug sales that impact consumption and how it reflects social media discussions is needed [71]. We limited this study to 8 broad drug categories due to the availability and abundance of related posts on the dark web; we hope to refine further and expand our categories for future work. Further, we

have identified processes for future research. We plan to expand this work to extract mental health symptoms from the drug-related social media data to connect the association between drugs and mental health problems, for example, the association between cannabis and depression [72,73]. We also plan to build an opioid drug social media knowledge graph (ODSM-KG) with all the diverse data points (drug, sentiment, emotion, mental health symptom, and location) and compare it against the state-of-the-art "Knowledge Graph based Approach For Exploring The U.S. Opioid Epidemic" [71]. Potential areas of application would be identifying risk factors regarding addiction and mental health from subreddit data [74], and identifying drug trends based on location with a possible opioid epidemic prediction. To identify the approaches for mitigating the misuse of opioids, it is imperative to study consumption patterns at the national and regional levels, the influences of the pharmaceutical industry, and the sociopolitical determinants that affect consumption. For our ODSM-KG, we aim to create a web-based tool that will allow for the depiction of historical patterns and enable comparisons between opioids, time periods, and areas within the United States. Given that our ODSM-KG was primarily based on data, we aim to enhance its accuracy and effectiveness by seeking guidance from a subject matter domain expert. This will enable us to customize it for unique scenarios and cater to the needs of specific users. We would also like to include Drug Enforcement Agency drug seizures in our preliminary data collection process to be aware of related social media discussions.

Global Relevance

SUD represents a substantial public health challenge worldwide, with far-reaching implications for individuals, families, and communities across diverse cultural and geographic contexts. Despite efforts by governments and health care organizations, the detection and intervention of SUDs remain complex and multifaceted, often hindered by stigma, limited resources, and barriers to early identification. We aim to examine the global relevance of detecting SUD through social media analysis, highlighting the potential of innovative approaches proposed in this study. Our study explores the potential of leveraging social media data, including content sourced from the dark web, to detect indicators of SUD and inform targeted interventions. By applying both temporal dynamics and domain-specific knowledge, we can extend this research to enhance the accuracy and effectiveness of SUD detection on a global scale. Collaborative efforts involving researchers, policy makers, and health care professionals from diverse regions can facilitate the adaptation and refinement of our approach to suit different cultural contexts and population groups. By harnessing the approaches implemented in this paper regarding temporal social media analysis and SUD-associated mental health disorders, researchers from other countries can adapt this approach to address these challenges associated with SUD and improve health outcomes for populations worldwide.

Acknowledgments

We acknowledge partial support from a National Institute on Drug Abuse (NIDA) award (1 R21DA044518: "eDarkTrends: Monitoring Cryptomarkets to Identify Emerging Trends of Illicit Synthetic Opioids Use") and National Science Foundation (NSF) awards (1761931: "Spokes: MEDIUM: MIDWEST: Collaborative: Community-Driven Data Engineering for Substance Abuse Prevention in the Rural Midwest" and NSF award: 2133842: "EAGER: Advancing Neuro-symbolic AI with Deep Knowledge-infused Learning"). The views or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the NSF and NIDA.

Conflicts of Interest

None declared.

References

1. Lamy FR, Daniulaityte R, Barratt MJ, Lokala U, Sheth A, Carlson RG. Listed for sale: analyzing data on fentanyl, fentanyl analogs and other novel synthetic opioids on one cryptomarket. Drug Alcohol Depend 2020 Aug 1;213:108115. [doi: 10.1016/j.drugalcdep.2020.108115] [Medline: 32585419]
2. Opioids. National Institute on Drug Abuse. URL: https://www.drugabuse.gov/drug-topics/opioids [accessed 2021-03-03]
3. Aldridge J, Décary-Hétu D. Hidden wholesale: the drug diffusing capacity of online drug cryptomarkets. Int J Drug Policy 2016 Sep;35:7-15. [doi: 10.1016/j.drugpo.2016.04.020] [Medline: 27260863]
4. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse. Trends in opioid use, harms, and treatment. In: Phillips JK, Ford MA, Bonnie RJ, editors. Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use: National Academies Press (US); 2017.
5. Global drug survey 2021. International Society of Substance Use Professionals. 2020 Dec 3. URL: https://www.issup.net/knowledge-share/news/2020-12/global-drug-survey-2021 [accessed 2021-03-03]
6. Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. Neuropsychopharmacology 2019 Feb;44(3):487-494. [doi: 10.1038/s41386-018-0247-x] [Medline: 30356094]

7. Liggett R, Lee JR, Roddy AL, Wallin MA. The dark web as a platform for crime: an exploration of illicit drug, firearm, CSAM, and cybercrime markets. In: Holt TJ, Bossler AM, editors. The Palgrave Handbook of International Cybercrime and Cyberdeviance: Palgrave Macmillan; 2020:91-116. [doi: 10.1007/978-3-319-78440-3]
8. Godawatte K, Raza M, Murtaza M, Saeed A. Dark web along with the dark web marketing and surveillance. Presented at: 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT); Dec 5 to 7, 2019; Gold Coast, QLD, Australia. [doi: 10.1109/PDCAT46702.2019.00095]
9. Branwen G, Christin N, Décary-Hétu D, Munksgaard R, Presidente E. Dark net market archives, 2011–2015. Gwern.net. 2015. URL: https://gwern.net/dnm-archive [accessed 2024-04-15]
10. ElBahrawy A, Alessandretti L, Rusnac L, Goldsmith D, Teytelboym A, Baronchelli A. Collective dynamics of dark web marketplaces. Sci Rep 2020 Nov 2;10(1):18827. [doi: 10.1038/s41598-020-74416-y] [Medline: 33139743]
11. Hayes DR, Cappa F, Cardon J. A framework for more effective dark web marketplace investigations. Information 2018;9(8):186. [doi: 10.3390/info9080186]
12. Harviainen JT, Haasio A, Hämäläinen L. Drug traders on a local dark web marketplace. Presented at: AcademicMindtrek ’20; Jan 29 to 30, 2020; Tampere, Finland URL: https://dl.acm.org/doi/proceedings/10.1145/3377290 [accessed 2024-04-23] [doi: 10.1145/3377290.3377293]
13. Haasio A, Harviainen JT, Savolainen R. Information needs of drug users on a local dark web marketplace. Inf Process Manag 2020 Mar;57(2):102080. [doi: 10.1016/j.ipm.2019.102080]
14. Lokala U, Lamy FR, Daniulaityte R, et al. Global trends, local harms: availability of fentanyl-type drugs on the dark web and accidental overdoses in Ohio. Comput Math Organ Theory 2019 Mar;25(1):48-59. [doi: 10.1007/s10588-018-09283-0] [Medline: 32577089]
15. Orsolini L, Papanti D, Corkery J, Schifano F. An insight into the deep web; why it matters for addiction psychiatry. Hum Psychopharmacol Clin Exp 2017 May;32(3):e2573. [doi: 10.1002/hup.2573] [Medline: 28657187]
16. Nguyen LT, Wu P, Chan W, Peng W, Zhang Y. Predicting collective sentiment dynamics from time-series social media. Presented at: WISDOM ’12: Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining; Aug 12 to 13, 2012; Beijing, China. [doi: 10.1145/2346676.2346682]
17. Iram, Aggarwal H. Time series analysis of PUBG and TikTok applications using sentiments obtained from social media-Twitter. Adv Math Sci J 2020;9(6):4047-4057. [doi: 10.37418/amsj.9.6.86]
18. Kolliakou A, Bakolis I, Chandran D, et al. Mental health-related conversations on social media and crisis episodes: a time-series regression analysis. Sci Rep 2020 Feb 6;10(1):1342. [doi: 10.1038/s41598-020-57835-9] [Medline: 32029754]
19. Pedersen MT, Andersen TO, Clotworthy A, et al. Time trends in mental health indicators during the initial 16 months of the COVID-19 pandemic in Denmark. BMC Psychiatry 2022 Jan 10;22(1):25. [doi: 10.1186/s12888-021-03655-8] [Medline: 35012486]
20. Obosi AC, Fatunbi AM, Oyinloye O. Peer pressure and substance use as predictors of mental health among in-school adolescents in Nigeria. Ianna J Interdiscip Stud 2022 Oct;4(1):1-9 [FREE Full text]
21. Ibrahim NF, Wang X. Decoding the sentiment dynamics of online retailing customers: time series analysis of social media. Comput Hum Behav 2019 Jul;96:32-45. [doi: 10.1016/j.chb.2019.02.004]
22. Hajiakhoond Bidoki N, Mantzaris AV, Sukthankar G. An LSTM model for predicting cross-platform bursts of social media activity. Information 2019;10(12):394. [doi: 10.3390/info10120394]
23. Sawhney R, Joshi H, Nobles A, Shah RR. Tweet classification to assist human moderation for suicide prevention. Proc Int AAAI Conf Weblogs Soc Media 2021 Jun 4;15:609-620. [Medline: 35173997]
24. Sawhney R, Joshi H, Gandhi S, Shah RR. A time-aware transformer based model for suicide ideation detection on social media. Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); Nov 16 to 20, 2020.. [doi: 10.18653/v1/2020.emnlp-main.619]
25. Lossio-Ventura JA, Bian J. An inside look at the opioid crisis over Twitter. Presented at: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 3 to 6, 2018; Madrid, Spain. [doi: 10.1109/BIBM.2018.8621101]
26. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of codeine misuse on Instagram: content analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22. [doi: 10.2196/publichealth.8144] [Medline: 29559422]
27. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling up research on drug abuse and addiction through social media big data. J Med Internet Res 2017 Oct 31;19(10):e353. [doi: 10.2196/jmir.6426] [Medline: 29089287]
28. Sarker A, Nataraj N, Siu W, Li S, Jones CM, Sumner SA. Concerns among people who use opioids during the COVID-19 pandemic: a natural language processing analysis of social media posts. Subst Abuse Treat Prev Policy 2022 Mar 5;17(1):16. [doi: 10.1186/s13011-022-00442-w] [Medline: 35248103]
29. Sarker A, Al-Garadi MA, Ge Y, Nataraj N, Jones CM, Sumner SA. Trends in co-mention of stimulants and opioids: a natural language processing driven analysis of Reddit forums. Res Square. Preprint posted online on Jan 17, 2022. [doi: 10.21203/rs.3.rs-1255278/v1]
30. Desrosiers A, Vine V, Kershaw T. “Ru mad?”: computerized text analysis of affect in social media relates to stress and substance use among ethnic minority emerging adult males. Anxiety Stress Coping 2019 Jan;32(1):109-123. [doi: 10.1080/10615806.2018.1539964] [Medline: 30373396]

31. Liu T, Giorgi S, Yadeta K, Schwartz HA, Ungar LH, Curtis B. Linguistic predictors from Facebook postings of substance use disorder treatment retention versus discontinuation. Am J Drug Alcohol Abuse 2022 Sep 3;48(5):573-585. [doi: 10.1080/00952990.2022.2091450] [Medline: 35853250]
32. Singh AK, Wu D. Sentiment analysis on substance use disorder (SUD) tweets before and during COVID-19 pandemic. In: Stephanidis C, Antona M, Ntoa S, editors. HCI International 2021 - Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part III: Springer; 2021:608-614. [doi: 10.1007/978-3-030-78645-8]
33. Cameron D, Smith GA, Daniulaityte R, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. J Biomed Inform 2013 Dec;46(6):985-997. [doi: 10.1016/j.jbi.2013.07.007] [Medline: 23892295]
34. Fan Y, Zhang Y, Ye Y, Li X, Zheng W. Social media for opioid addiction epidemiology: automatic detection of opioid addicts from Twitter and case studies. Presented at: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management; Nov 6 to 10, 2017; Singapore, Singapore. [doi: 10.1145/3132847.3132857]
35. Eshleman R, Jha D, Singh R. Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media. Presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov 13 to 16, 2017; Kansas City, MO. [doi: 10.1109/BIBM.2017.8217766]
36. Khandelwal S. Dark web users suspect "Dream Market" has also been backdoored by feds. The Hacker News. 2017 Jul 21. URL: https://thehackernews.com/2017/07/dream-market-darkweb.html [accessed 2022-10-11]
37. Whittaker Z. Deep dot web indictment. DocumentCloud. 2019 Apr 24. URL: https://www.documentcloud.org/documents/5993699-Deep-Dot-Web-Indictment.html [accessed 2024-03-26]
38. Smith K. Tochka market Url: links- buy items from Tochka Darknet marketplace. Deep Web. 2016. URL: https://www.deepweb-sites.com/tochka-market-url-linksdarknet-reddit-review/ [accessed 2022-10-14]
39. Dark web marketplace Wall Street market busted by international police. Sophos News. 2019. URL: https://nakedsecurity.sophos.com/2019/05/07/dark-web-marketplace-wallstreet-market-busted-by-international-police/ [accessed 2023-11-12]
40. van Wirdum A. Major darknet marketplace Wall Street market shuttered by law. Bitcoin Magazine. 2019 May 3. URL: https://bitcoinmagazine.com/culture/major-darknet-marketplace-wall-street-market-shuttered-law-enforcement [accessed 2024-03-26]
41. Greenberg A. Drug market 'Agora' replaces the Silk Road as king of the dark net. WIRED. 2014 Sep 2. URL: https://www.wired.com/2014/09/agora-bigger-than-silk-road/ [accessed 2022-10-11]
42. Chiu JPC, Nichols E. Named entity recognition with bidirectional LSTM-CNNs. Trans Assoc Computational Linguistics 2016 Dec;4:357-370. [doi: 10.1162/tacl_a_00104]
43. Lokala U, Lamy F, Daniulaityte R, et al. Drug abuse ontology to harness web-based data for substance use epidemiology research: ontology development study. JMIR Public Health Surveill 2022 Dec 23;8(12):e24938. [doi: 10.2196/24938] [Medline: 36563032]
44. García J, García-Peñalvo FJ, Therón R. A survey on ontology metrics. In: Lytras MD, De Pablos PO, Ziderman A, Roulstone A, Maurer H, Imber JB, editors. Knowledge Management, Information Systems, E-Learning, and Sustainability Research: Third World Summit on the Knowledge Society, WSKS 2010, Corfu, Greece, September 22-24, 2010, Proceedings, Part I: Springer; 2010:22-27. [doi: 10.1007/978-3-642-16318-0]
45. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online on Mar 11, 2022. [doi: 10.48550/arXiv.2203.05794]
46. Hussain S, Ahmad S, Saleem M, Akhtar S. Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. PLoS One 2020 Sep 28;15(9):e0239098. [doi: 10.1371/journal.pone.0239098] [Medline: 32986764]
47. Biswas S, Ghosh S. Drug usage analysis by VADER sentiment analysis on leading countries. Mapana J Sci 2022;21(3):1. [doi: 10.12723/mjs.62.0]
48. Wang W, Chen L, Thirunarayan K, Sheth AP. Harnessing Twitter “big data” for automatic emotion identification. Presented at: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing; Sep 3 to 5, 2012; Amsterdam, Netherlands. [doi: 10.1109/SocialCom-PASSAT.2012.119]
49. Mozafari M, Farahbakhsh R, Crespi N. A BERT-based transfer learning approach for hate speech detection in online social media. In: Cherifi H, Gaito S, Mendes JF, Moro E, Rocha LM, editors. Complex Networks and Their Applications VIII: Volume 1 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019: Springer; 2019:928-940. [doi: 10.1007/978-3-030-36687-2]
50. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S. Semeval-2018 task 1: affect in tweets. Presented at: Proceedings of The 12th International Workshop on Semantic Evaluation; Jun 5 to 6, 2018; New Orleans, LA URL: http://aclweb.org/anthology/S18-1 [accessed 2024-04-23] [doi: 10.18653/v1/S18-1001]
51. Lokala U, Daniulaityte R, Carlson R, Lamy F, Yadav S, Sheth A. Social media data for exploring the association between cannabis use and depression. Figshare. 2021 Feb 25. URL: https://doi.org/10.6084/m9.figshare.14067122.v3 [accessed 2024-04-17]
52. Wang Y, Yang N, Miao D, Chen Q. Dual-channel and multigranularity gated graph attention network for aspect-based sentiment analysis. Appl Intell 2022 Oct 6;53:13145-13157. [doi: 10.1007/s10489-022-04198-5]

53. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv. Preprint posted online on Dec 22, 2014. [doi: 10.48550/arXiv.1412.6980]
54. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv. Preprint posted online on Jul 3, 2012. [doi: 10.48550/arXiv.1207.0580]
55. Ramkumar S, Sariki TP, Kumar GB, Kannan RJ. Detecting suicidal ideation from online texts. In: Kannan RJ, Geetha S, Sashikumar S, Diver C, editors. International Virtual Conference on Industry 40: Select Proceedings of IVCI40 2020: Springer; 2021:413-425. [doi: 10.1007/978-981-16-1244-2]
56. Ji S, Yu CP, Fung S, Pan S, Long G. Supervised learning for suicidal ideation detection in online user content. Complexity 2018 Sep 9;2018:1-10. [doi: 10.1155/2018/6157249]
57. Pitsilis GK, Ramampiaro H, Langseth H. Effective hate-speech detection in Twitter data using recurrent neural networks. Appl Intell 2018 Jul 26;48(12):4730-4742. [doi: 10.1007/s10489-018-1242-y]
58. Cheng JC, Chen ALP. Multimodal time-aware attention networks for depression detection. J Intell Inf Syst 2022 Oct;59:319-339. [doi: 10.1007/s10844-022-00704-w]
59. Henderson M, Johnson NF, Auld G. Silences of ethical practice: dilemmas for researchers using social media. Educ Res Eval 2013 Jun 28;19(6):546-560. [doi: 10.1080/13803611.2013.805656]
60. Alim S. An initial exploration of ethical research practices regarding automated data extraction from online social media user profiles. First Monday 2014 Jul 6;19(7). [doi: 10.5210/fm.v19i7.5382]
61. Matthews T, O'Leary K, Turner A, et al. Stories from survivors: privacy & security practices when coping with intimate partner abuse. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6 to 11, 2017; Denver, CO. [doi: 10.1145/3025453.3025875]
62. Hunter RF, Gough A, O'Kane N, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
63. Bruckman A. Studying the amateur artist: a perspective on disguising data collected in human subjects research on the internet. Ethics Inf Technol 2002 Sep;4:217-231. [doi: 10.1023/A:1021316409277]
64. Preoţiuc-Pietro D, Volkova S, Lampos V, Bachrach Y, Aletras N. Studying user income through language, behaviour and affect in social media. PLoS One 2015 Sep 12;10(9):e0138717. [doi: 10.1371/journal.pone.0138717] [Medline: 26394145]
65. Wang Z, Hale S, Adelani DI, et al. Demographic inference and representative population estimates from multilingual social media data. Presented at: WWW '19: The World Wide Web Conference; May 13 to Mar 17, 2019; San Francisco, CA. [doi: 10.1145/3308558.3313684]
66. Kaserer T, Steinacher T, Kainhofer R, et al. Identification and characterization of plant-derived alkaloids, corydine and corydaline, as novel mu opioid receptor agonists. Sci Rep 2020 Aug 14;10(1):13804. [doi: 10.1038/s41598-020-70493-1] [Medline: 32796875]
67. Just J, Mücke M, Bleckwenn M. Dependence on prescription opioids: prevention, diagnosis and treatment. Deutsches Arztebl Int 2016;113(13):213-220. [doi: 10.3238/arztebl.2016.0213]
68. Landry MJ, Smith DE, McDuff DR, Baughman OL. Benzodiazepine dependence and withdrawal: identification and medical management. J Am Board Fam Pract Mar-Apr 1992;5(2):167-175. [Medline: 1575069]
69. Woolson RF. Wilcoxon signed-rank test. In: Encyclopedia of Biostatistics 2005. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/0470011815 [accessed 2024-04-15] [doi: 10.1002/0470011815]
70. Kumar R, Yadav S, Daniulaityte R, et al. eDarkFind: unsupervised multi-view learning for Sybil account detection. Presented at: WWW '20; Apr 20 to 24, 2020; Taipei, Taiwan URL: https://dl.acm.org/doi/proceedings/10.1145/3366423 [accessed 2024-04-23] [doi: 10.1145/3366423.3380263]
71. Kamdar MR, Hamamsy T, Shelton S, et al. A knowledge graph-based approach for exploring the U.S. opioid epidemic. arXiv. Preprint posted online on May 27, 2019. [doi: 10.48550/arXiv.1905.11513]
72. Roy K, Lokala U, Khandelwal V, Sheth A. "Is depression related to cannabis?": a knowledge-infused model for entity and relation extraction with limited supervision. arXiv. Preprint posted online on Feb 1, 2021. [doi: 10.48550/arXiv.2102.01222]
73. Yadav S, Lokala U, Daniulaityte R, Thirunarayan K, Lamy F, Sheth A. "When they say weed causes depression, but it's your fav antidepressant": knowledge-aware attention framework for relationship extraction. PLoS One 2021 Mar 25;16(3):e0248299. [doi: 10.1371/journal.pone.0248299] [Medline: 33764983]
74. Gaur M, Kursuncu U, Alambo A, et al. "Let me tell you about your mental health!": contextualized classification of Reddit posts to DSM-5 for web-based intervention. Presented at: CIKM '18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management; Oct 22 to 26, 2018; Torino, Italy.

Abbreviations

BERT: Bidirectional Encoder Representations From Transformers
CNN: convolutional neural network
DAO: drug abuse ontology
H-LSTM: history-aware long short-term memory
H-RNN: history-aware recurrent neural network

LIWC: Linguistic Inquiry and Word Count
LSTM: long short-term memory
NER: named-entity recognition
ODSM-KG: opioid drug social media knowledge graph
POS: part of speech
PREDOSE: Prescription Drug Abuse Online Surveillance and Epidemiology
SUD: substance use disorder
SUDA: substance use disorder absent
SUDP: substance use disorder present
SUDS: Substance Use Disorder Corpus
TF-IDF: term frequency–inverse document frequency
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by CN Hang; submitted 26.04.23; peer-reviewed by Anonymous, Anonymous, V Subramaniyan; revised version received 16.02.24; accepted 21.02.24; published 01.05.24.

Please cite as:
Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A
Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study
JMIRx Med 2024;5:e48519
URL: https://xmed.jmir.org/2024/1/e48519
doi:10.2196/48519

Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis

Tim Dong[1], MSc; Shubhra Sinha[1], MBBS; Ben Zhai[2], PhD; Daniel Fudulu[1], MD, PhD; Jeremy Chan[1], MD; Pradeep Narayan[3], MD; Andy Judge[1], PhD; Massimo Caputo[1], MD; Arnaldo Dimagli[1], MD; Umberto Benedetto[1], MD, PhD; Gianni D Angelini[1], MD

1

2

3

Corresponding Author:
Tim Dong, MSc

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45973

Companion article: https://www.medrxiv.org/content/10.1101/2023.01.21.23284795v1

Companion article: https://med.jmirx.org/2024/1/e60428

Companion article: https://med.jmirx.org/2024/1/e60280

Companion article: https://med.jmirx.org/2024/1/e60384

Abstract

Background: The Society of Thoracic Surgeons and European System for Cardiac Operative Risk Evaluation (EuroSCORE) II risk scores are the most commonly used risk prediction models for in-hospital mortality after adult cardiac surgery. However, they are prone to miscalibration over time and poor generalization across data sets; thus, their use remains controversial. Despite increased interest, a gap in understanding the effect of data set drift on the performance of machine learning (ML) over time remains a barrier to its wider use in clinical practice. Data set drift occurs when an ML system underperforms because of a mismatch between the data it was developed from and the data on which it is deployed.

Objective: In this study, we analyzed the extent of performance drift using models built on a large UK cardiac surgery database. The objectives were to (1) rank and assess the extent of performance drift in cardiac surgery risk ML models over time and (2) investigate any potential influence of data set drift and variable importance drift on performance drift.

Methods: We conducted a retrospective analysis of prospectively, routinely gathered data on adult patients undergoing cardiac surgery in the United Kingdom between 2012 and 2019. We temporally split the data 70:30 into a training and validation set and a holdout set. Five novel ML mortality prediction models were developed and assessed, along with EuroSCORE II, for relationships between and within variable importance drift, performance drift, and actual data set drift. Performance was assessed using a consensus metric.

Results: A total of 227,087 adults underwent cardiac surgery during the study period, with a mortality rate of 2.76% (n=6258). There was strong evidence of a decrease in overall performance across all models (P<.0001). Extreme gradient boosting (clinical effectiveness metric [CEM] 0.728, 95% CI 0.728-0.729) and random forest (CEM 0.727, 95% CI 0.727-0.728) were the overall best-performing models, both temporally and nontemporally. EuroSCORE II performed the worst across all comparisons. Sharp changes in variable importance and data set drift from October to December 2017, from June to July 2018, and from December 2018 to February 2019 mirrored the effects of performance decrease across models.

Conclusions: All models show a decrease in at least 3 of the 5 individual metrics. CEM and variable importance drift detection demonstrate the limitation of logistic regression methods used for cardiac surgery risk prediction and the effects of data set drift. Future work will be required to determine the interplay between ML models and whether ensemble models could improve on their respective performance advantages.

(JMIRx Med 2024;5:e45973) doi:10.2196/45973

KEYWORDS

cardiac surgery; artificial intelligence; risk prediction; machine learning; operative mortality; data set drift; performance drift; national data set; adult; data; cardiac; surgery; cardiology; heart; risk; prediction; United Kingdom; mortality; performance; model

Introduction

Background

Recently, the importance of machine learning (ML), a branch of artificial intelligence, has been highlighted as a potential alternative to traditional mortality risk stratification models such as the Society of Thoracic Surgeons (STS) [1] and European System for Cardiac Operative Risk Evaluation (EuroSCORE) II risk scores [2], which are prone to miscalibration over time and poor generalization across data sets [1,3]. These traditional scoring methods are generally based on logistic regression (LR), with risk factors determined through consensus across experts within leading cardiac surgery organizations in the United States (STS) or Europe (EuroSCORE II). In particular, EuroSCORE II, which is based on LR using 18 items of information about the patient, has been shown by numerous studies to display poor discrimination and calibration across data sets with differing characteristics, including but not limited to age [4], ethnicity [5], and procedures groups [6-10].

Risk scoring models' performance is challenged by numerous factors, such as differences in variable definitions, the management of incomplete data fields, surgical procedure selection criteria, and temporal changes in the prevalence of patients' risk factors [11]. ML approaches are increasingly used for prediction in health care research as they have the potential to overcome the limitations of linear models. By including pairwise and higher-order interactions and modeling nonlinear effects, ML may overcome heterogeneity in procedures and missing data [1,12]. Although ML has been shown to be beneficial over conventional scoring systems, the magnitude and clinical influence of such improvements remain uncertain [2]. The ability to counter "performance drift" due to temporal changes in the prevalence of risk factors has also yet to be fully elucidated.

In ML, performance drift refers to the gradual loss in model performance caused by changes that call into question the model's training assumptions. Key causes of performance drift include data set drift, which refers to changes in the distribution of data between training and evaluation sets; variable importance drift, which involves changes in the significance of model variables; and calibration drift, which is characterized by decreased reliability in estimated probabilities. These factors can interact, as seen in a study of noncardiac surgery [13]. Understanding the complex relationship between variable importance drift, performance drift, and data set drift is important. This relationship explains how changes in the importance of specific variables, combined with changes in the actual data distribution, collectively influence the model's overall accuracy and reliability as it performs over time. The wider implications are also significant, influencing decision-making, insight accuracy, generalization [14], ethical considerations, and regulatory compliance across industries.

The aim of this study was to investigate performance drift in existing ML models that have been used in prior cardiac surgery risk prediction research. The objectives were to (1) rank and assess the extent of performance drift in such cardiac surgery risk ML models over time and (2) investigate any potential influence of data set drift and variable importance drift on performance drift. Therefore, we trained and evaluated 5 supervised ML models in addition to EuroSCORE II to (1) determine the best ML model in terms of overall accuracy, discrimination, calibration, and clinical effectiveness; (2) use variable importance drift as a measure for detecting data set drift; and (3) verify suspected data set drift informed through variable importance drift by assessing actual data set drift [15].

Related Work

In our previous study, we found that combining the metrics covering all 4 aspects of discrimination, calibration, clinical usefulness, and overall accuracy into a single clinical effectiveness metric (CEM) improved the efficiency of cognitive decision-making (according to the Miller law [16]) for selecting the optimal ensemble models (ie, using several models to derive a consensus prediction) [14,17]. This approach is useful for providing a consensus metric that enables models to be ranked in scenarios where, for example, 1 model could outperform another using 1 metric but underperform under a different metric. Furthermore, we demonstrated that such a consensus metric could be combined with drill-down analysis to further interpret the models using individual metrics [14]. Although area under the curve (AUC) evaluates the diagnostic or predictive performance of the model, it does not directly reflect patient benefit. This is why we included a suit of other metrics, including the decision curve analysis (DCA) net benefit index, that were found to be clinically pertinent from our prior study [18].

In our previous work [19], we studied the calibration changes across 2 different time intervals using the calibration belt (overall external calibration) and calibration drift (Hosmer-Lemeshow goodness-of-fit χ^2 statistics) approaches within a single UK hospital. A recent study extended our work to a Chinese national registry, Sino (Chinese) System for Coronary Artery Bypass Grafting (CABG) Operative Risk Evaluation II (SinoSCORE II), using a set of ML models such as LightGBM; CatBoost; and a combination of variable selection approaches including Optuna for stepwise regression, BorutaSHAP, and feature importance ranking [20]. Another study in the United States also investigated the calibration performance difference between extreme gradient boosting (XGBoost) and LR models built for a cohort of patients who underwent CABG, using preoperative, intraoperative, and combined variable sets from the STS Adult Cardiac Surgery Database [21].

Methods

Data Set and Patient Population

The study was performed using the National Adult Cardiac Surgery Audit (NACSA) data set, which comprises data prospectively collected by the National Institute for Cardiovascular Outcomes Research on all cardiac procedures performed in all National Health Service hospitals and some private hospitals across the United Kingdom [19].

A total of 227,087 adult patients who underwent cardiac surgery between January 1, 2012, and March 31, 2019, were included. Congenital, transplant, and mechanical support device insertion cases were excluded. The CONSORT (Consolidated Standards of Reporting Trials) patient flow diagram is shown in Figure S1 in Multimedia Appendix 1 [19,22-25]. Missing and erroneously inputted data in the data set were cleaned according to the NACSA registry data preprocessing recommendations [26]. Generally, for any variable data that were missing, it was assumed that the variable was at baseline level, that is, no risk factor was present. Missing patient age at the time of surgery was imputed as the median patient age for the corresponding year. Data standardization was performed by subtracting the variable mean and dividing by the SD values [22].

The data set was split into 2 cohorts: training and validation set (n=157,196, 69.2%; 2012-2016; Table S1 in Multimedia Appendix 1) and holdout set (n=69,891, 30.8%; 2017-2019; Table S2 in Multimedia Appendix 1). The primary outcome of this study was in-hospital mortality.

Baseline Statistical Analysis

Continuous variables were compared using nonparametric Wilcoxon rank sum tests, whereas categorical variables were compared using Pearson χ^2 tests or Fisher exact tests as appropriate.

The *Scikit-learn* (version 0.23.1) and *Keras* (version 2.4.0) Python libraries (Python Software Foundation) were used to develop the models and to evaluate their discrimination, calibration, and clinical effectiveness capabilities. Statistical analyses were conducted using Stata/MP (version 17; StataCorp) and R (version 4.0.2; StataCorp). ANOVA assumptions were checked using the *rstatix* R package.

Model Development

In our study, we trained 5 supervised ML risk models based on the EuroSCORE II preoperative variable set (Table S3 in Multimedia Appendix 1). Those 5 models included LR, neural network (NN) [22], random forest (RF) [27], weighted support vector machine (SVM) [28], and extreme gradient boosting (XGBoost) [19,29]. The EuroSCORE II score was calculated for baseline comparison. Internal validation was performed using 5-fold cross-validation on the training and validation set (2012-2016) to select model parameters. Final models were determined by retraining the models on the combined training and validation set using the selected model parameters. Temporal external validation was performed using the final models on the holdout set (2017-2019) [15]. Each model calculated the probability of surgical mortality for each patient. Overall, 1000 bootstrap samples were taken for all metrics. Further details on model development can be found in the *Model Specification* section in Multimedia Appendix 1.

Assessment of Model Performance

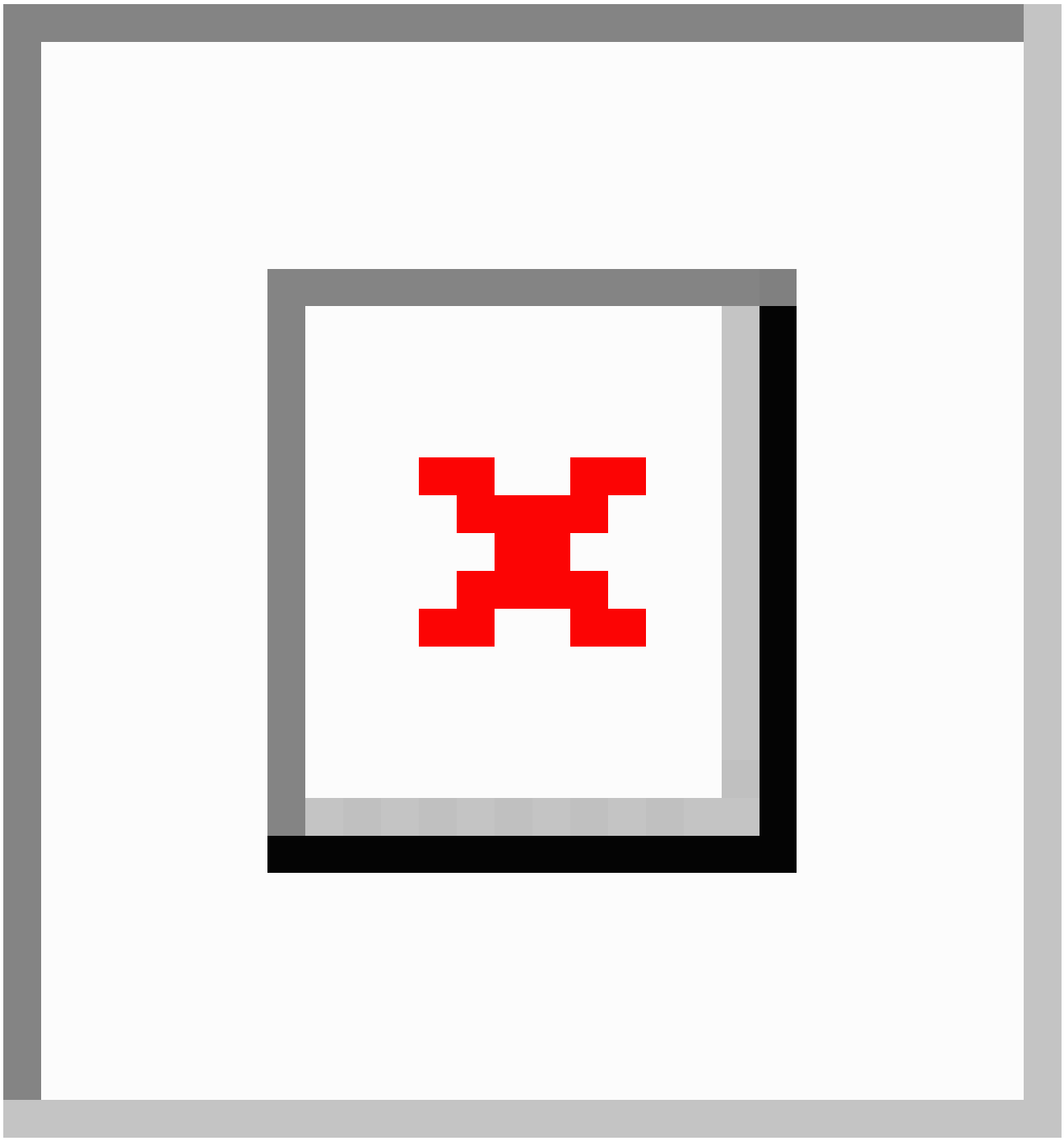
The models' performance was measured across four broad parameters:

1. Discrimination: AUC and F_1-score
2. Clinical utility: DCA net benefit index
3. Calibration: 1 – expected calibration error (ECE)
4. Combination of calibration and discrimination: adjusted Brier score

The AUC performances of all variant models were evaluated, and the receiver operating characteristic (ROC) curves were plotted [30]. As a sensitivity analysis, we calculated the F_1-score, which combines precision and recall without explicitly considering the true negative rate in the performance evaluation [31]. This metric adjusts for the biased effect due to the high proportion of alive outcome samples. The DCA net benefit index was used to test clinical benefit [32]. 1 – ECE was used to determine calibration performance, with higher values being better [33]. A special case of the Brier score (1 – Brier score) without the normalization term was used (adjusted Brier score) [34], with higher values indicating better discrimination and calibration performance.

To determine the best model in terms of both discrimination and calibration, we took the geometric average of AUC, F_1-score [31], DCA net benefit index (treated + untreated), 1 – ECE, and 1 – Brier score. The consensus metric using the combined geometric average of the 5 metrics is named CEM for ease of reference. The consensus approach for combining different metrics has previously been applied in a study on COVID-19 prediction [35]. In addition, this approach is similar to the simple additive weighting multicriteria evaluation approach for making a decision through the ranking of a set of competing criteria [36]. Geometric average has previously been found to be effective for summarizing metrics for temporal-based model calibration and is robust for bootstrap-sampled Gaussian distributions [37]. This metric is robust to outliers [38] and is preferable for aggregation compared to the weighted arithmetic mean [39]. As an exception, the arithmetic average was used for the DCA net benefit index over all thresholds as a measure of overall net benefit, before geometric averaging, since the values can be negative. An overview of the model and evaluation design is shown in Figure 1.

Figure 1. Design overview of the study. Nontemporal performance and drift (temporal) analyses were performed. Drifts in discrimination, calibration, clinical utility, data set, and variable importance were assessed. Time point assessments were performed for the clinical effectiveness metric (CEM). Drifts in component metrics of CEM were evaluated. AUC: area under the curve; ECE: expected calibration error; EuroSCORE: European System for Cardiac Operative Risk Evaluation; F1: F_1-score; neuronetwork: neural network; SVM: support vector machine; Xgboost: extreme gradient boosting.



Baseline Nontemporal Performance

Nontemporal comparison of models was conducted as a baseline, using all data across the holdout period. Differences across models were tested using repeated-measures 1-way ANOVA and Bonferroni-corrected, multiple pairwise, paired *t* tests (1-tailed); this was followed by Dunnett correction for multiple comparisons, with the overall best-performing model as the control. ANOVA assumptions for outliers were checked. Normality assumptions were checked using the Shapiro-Wilk test [40]. The Delong test was applied to determine whether there was a statistically significant difference across the AUCs of the ROC curves for the top 2 best-performing models. A comparison of individual metrics was conducted.

Drift Analysis

Overview

The statistical methods used for analyzing drift is shown in Table 1. More detailed explanations are provided below.

Table . Summary of statistical methods used for assessing drift.

Objective and statistical tests		General statistical situations	Rationale for choosing test	Assumptions checked
Nontemporal comparison of models				
	Repeated-measures 1-way ANOVA	Comparison of multiple groups for differences	Used for comparing means across multiple models	Outliers (ANOVA assumptions) and normality (Shapiro-Wilk test)
	Paired *t* tests (Bonferroni corrected)	Comparison of paired observations between models	To compare specific model pairs simultaneously	—[a]
	Dunnett correction	Control for multiple comparisons	Controls type I error rate in comparing multiple treatments to a control group in 1-way ANOVA	—
	Delong test	Comparison of the AUCs[b] of 2 correlated ROC[c] curves	To compare the AUCs of 2 models or tests during sensitivity testing	—
Analysis within specific time frames				
	Kruskal-Wallis Test	Comparison of multiple groups for differences (nonparametric)	Nonparametric alternative for ANOVA in specific time frames	Outliers (ANOVA assumptions) and normality (Shapiro-Wilk test)
	Bonferroni-corrected, paired-samples Wilcoxon test (Wilcoxon signed rank test)	Comparison of paired observations within time frames	Nonparametric comparison of paired samples within time frames, with control for type I error rate in comparing multiple treatments	—
	Dunn test	Multiple pairwise comparisons within nonparametric groups	Post hoc test for pairwise comparisons after Kruskal-Wallis test; determines the magnitude of difference effects within time frames	—
Analysis between the first 3 months of 2017 and 2019				
	Kruskal-Wallis test	Comparison of multiple groups for differences (nonparametric)	Nonparametric comparison between time frames	Outliers (ANOVA assumptions) and normality (Kolmogorov-Smirnov Test)
	Paired-samples Wilcoxon test (Wilcoxon signed rank test)	Comparison of paired observations between time frames	Nonparametric comparison of paired samples between time frames	—
	Bonferroni-adjusted Dunn test	Multiple pairwise comparisons between time frames	Post hoc test for pairwise comparisons after significant Kruskal-Wallis results; determines the magnitude of difference effects between time frames, with control for type I error rate in comparing multiple treatments	Normality (Kolmogorov-Smirnov Test)
Analysis of discrimination, calibration, clinical utility, and overall accuracy drift				
	Linear regression (with residual analysis)	Assessing relationships and regression parameters	To analyze linear relationships and model residuals	Normality through histograms and QQ plots
	Seasonal Kendall test (nonparametric alternative if assumptions not met)	Assessing association or trends when assumptions are not met	Nonparametric test for assessing associations without assumptions	Homoscedasticity through scale-location plots

[a]Not applicable.

[b]AUC: area under the curve.

[c]ROC: receiver operating characteristic.

CEM Regression Trends

The geometric CEM mean (and 95% CI) value of 1000 bootstraps for each model against time (the month of the year) was calculated, and the results were plotted to compare trends across models. The models were compared by fitting multiple linear regression lines across time for CEM.

To check for normality assumptions, we plotted the histogram and a QQ plot of residuals before applying linear regressions [41]. We also checked for homogeneity of residual variance (homoscedasticity) by plotting a scale-location plot, that is, the square root of standardized residual points against the values of the fitted outcome variable [42]. For model metrics that do not satisfy these assumptions, the seasonal Kendall test (nonparametric) was used instead.

Analysis Within the First 3 Months of 2017 and 2019

Differences in CEM values across models at 2 time points were independently tested using the Kruskal-Wallis test and Bonferroni-corrected, paired-samples Wilcoxon test (Wilcoxon signed rank test). The 2 time points were the first 3 months of 2017 and 2019. This was followed by the Dunn test for nonparametric multiple comparisons of the models at each of the 2 time points, with the overall best-performing model as a baseline. ANOVA assumptions for outliers were checked. Normality assumptions were checked using the Shapiro-Wilk test [40].

Analysis Between the First 3 Months of 2017 and 2019

Differences in CEM values across the first 3 months of 2017 and 2019 were tested using the Kruskal-Wallis test and paired-samples Wilcoxon test (Wilcoxon signed rank test). The Bonferroni-adjusted Dunn test was used to determine the magnitude and evidence of change across the 2 time points for each model. ANOVA assumptions for outliers were checked. Normality assumptions were checked using the Kolmogorov-Smirnov Test.

Analysis of Discrimination, Calibration, Clinical Utility, and Overall Accuracy Drift

As a sensitivity analysis, we analyzed performance drift in terms of component metrics within CEM. Discrimination (AUC), positive outcome discrimination (F_1-score), calibration (1 – ECE), clinical utility (net benefit), and overall accuracy of prediction probability (adjusted Brier score) were assessed by fitting multiple (model) linear regression lines across time for each metric.

To check for normality assumptions, the same methods as those used for CEM regression trends were used.

Analysis of Variable Importance Drift

Variable importance drift was assessed for the best-performing model. For each month of the holdout set, 5-fold nested cross-validation was performed to derive the importance of each EuroSCORE II variable in the model's decision-making. The geometric mean of 5-fold importance at each time point was plotted along with the importance of each of the 5 folds. The Shapley additive explanations (SHAP) mean absolute magnitude of importance was used [43,44]. Locally estimated scatterplot smoothing was used to simplify the visual representation. Line plots of the top 6 most important variables were used as a sensitivity analysis.

Data Set Drift

Data set drift across time was visualized using a stacked bar plot for the top 3 variables as identified by SHAP variable importance. Continuous variables were binned into intervals to enable ease of analysis.

Net Benefit Projection

To further understand the clinical significance of the performance drift over time, the fitted linear regression model intercepts and slopes were used to extrapolate the net benefit up to January 2030 for the XGBoost and NN models.

Ethical Considerations

The study was part of a research project approved by the Health Research Authority and Health and Care Research Wales on July 23, 2019 (Integrated Research Application System project ID: 257758). As the study included retrospective interrogation of the National Institute for Cardiovascular Outcomes Research database, the need for individual patient consent was waived in accordance with the research guidance. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Results

Baseline Patient Characteristics

A total of 227,087 procedures of adults from 42 hospitals were included in this analysis. This followed the removal of 3930 congenital cases, 1586 transplant and mechanical support device insertion cases, and 3395 procedures with missing information on mortality (Table 2). There were 6258 deaths during the study period (mortality rate of 2.76%).

Table . Patient demographics and summary of cleaned EuroSCORE[a] II variables. Variables are from the time period from 2012 to 2019. Records with missing mortality status were excluded.

Variable		Mortality status		*P* value[b]
		No (n=220,829)	Yes (n=6258)	
Age (years), mean (SD)		67.53 (11.23)	70.77 (11.42)	<.001
NYHA[c] classification, n (%)				<.001
	0 (I)	48,625 (22)	1055 (17)	
	1 (II)	96,888 (44)	1609 (26)	
	2 (III)	64,049 (29)	2228 (36)	
	3 (IV)	11,267 (5.1)	1366 (22)	
Renal impairment, n (%)				<.001
	0 (normal)	103,196 (47)	1704 (27)	
	1 (moderate)	92,411 (42)	2451 (39)	
	2 (on dialysis)	2187 (1)	330 (5.3)	
	3 (severe)	23,035 (10)	1773 (28)	
Chronic lung disease, n (%)		26,644 (12)	1211 (19)	<.001
Poor mobility, n (%)		8305 (3.8)	514 (8.2)	<.001
Previous cardiac surgery, n (%)		12,012 (5.4)	1141 (18)	<.001
Left ventricle function, n (%)				<.001
	0 (good; >50%)	184,721 (84)	4706 (75)	
	1 (moderate; 31%-50%)	30,608 (14)	1089 (17)	
	2 (poor; 21%-30%)	4241 (1.9)	318 (5.1)	
	3 (very poor; ≤20%)	1259 (0.6)	145 (2.3)	
Pulmonary hypertension, n (%)				<.001
	0 (PA[d] systolic <31 mm Hg)	201,643 (91)	5000 (80)	
	1 (PA systolic 31-55 mm Hg)	13,126 (5.9)	705 (11)	
	2 (PA systolic >55 mm Hg)	6060 (2.7)	553 (8.8)	
CCS[e] class 4 angina, n (%)		18,370 (8.3)	956 (15)	<.001
Urgency, n (%)				<.001
	0 (elective)	141,617 (64)	2442 (39)	
	1 (urgent)	72,090 (33)	2134 (34)	
	2 (emergency)	6533 (3)	1230 (20)	
	3 (salvage)	589 (0.3)	452 (7.2)	
Weight of the intervention, n (%)				<.001
	0 (isolated CABG[f])	111,243 (50)	1546 (25)	
	1 (single non-CABG)	62,568 (28)	2153 (34)	
	2 (two procedures)	42,649 (19)	2108 (34)	
	3 (three procedures)	4369 (2)	451 (7.2)	
Diabetes on insulin, n (%)		12,818 (5.8)	453 (7.2)	<.001
Female gender, n (%)		59,467 (27)	2328 (37)	<.001
Recent myocardial infarction, n (%)		43,316 (20)	1594 (25)	<.001
Critical preoperative state, n (%)		7255 (3.3)	1382 (22)	<.001

Variable	Mortality status		*P* value[b]
	No (n=220,829)	Yes (n=6258)	
Extracardiac arteriopathy, n (%)	22,327 (10)	1215 (19)	<.001
Active endocarditis, n (%)	5816 (2.6)	493 (7.9)	<.001
Surgery on thoracic aorta, n (%)	9070 (4.1)	896 (14)	<.001
EuroSCORE II, mean (SD)	0.03 (0.04)	0.12 (0.14)	<.001

[a]EuroSCORE: European System for Cardiac Operative Risk Evaluation.

[b]Wilcoxon rank sum test or Pearson χ^2 test

[c]NYHA: New York Heart Association.

[d]PA: pulmonary artery.

[e]CCS: Canadian Cardiovascular Society.

[f]CABG: coronary artery bypass grafting.

Baseline Nontemporal Performance

No extreme outliers were found when testing for ANOVA assumptions. The CEM values from 1000 bootstraps were normally distributed for LR, NN, and RF but not XGBoost, as assessed by the Shapiro-Wilk test (*P*>.05). A histogram plot of the XGBoost CEM values did not show substantial deviation from the normal distribution. There was strong evidence of a difference across all models (*P*<.0001; Table S4 and Figure S2 in Multimedia Appendix 1). Table 3 shows that XGBoost (CEM 0.728, 95% CI 0.728-0.729) and RF (CEM 0.727, 95% CI 0.727-0.728) were the overall best-performing models, with moderate to strong evidence (nonoverlapping CIs) of the former outperforming the latter. This was followed by LR, NN, SVM, and EuroSCORE II. The Dunnett test showed that there was moderate to strong evidence that XGBoost was superior to all other models (*P*<.001; Table 4). The performance of XGBoost was the least different from RF but the most different from EuroSCORE II (CEM difference to XGBoost: 0.0009 vs 0.1876).

The sensitivity analysis of CEM component metrics showed that the adjusted Brier score was unable to distinguish between XGBoost, RF, NN, and LR (Table 3; all 0.976). AUC performance was the best for XGBoost (0.834) and RF (0.835), with the Delong test showing no statistically significant difference (*P*>.05). F_1-score showed that XGBoost performed the best, followed by RF (0.279 vs 0.277). LR and NN (adjusted ECE: both 0.997) showed better calibration performance than RF and XGBoost (adjusted ECE: both 0.996). Net benefit was the best for XGBoost and RF (both 0.904).

Table . Geometric mean of individual metrics for each model in the holdout set. In all, 1000 bootstrap samples were used to derive the geometric mean of each metric. Adjusted ECE[a] and Brier score values are shown. Net benefit is the average absolute overall benefit across all thresholds.

Model category	1 – ECE	AUC[b]	1 – Brier score	F_1-score	Net benefit	CEM[c]		
						Mean (SD)	95% CI	Value, n
EuroSCORE[d] II	0.641	0.800	0.814	0.240	0.461	0.541 (0.004)	0.540-0.541	1000
LR[e]	0.997	0.819	0.976	0.264	0.902	0.717 (0.005)	0.717-0.717	1000
NN[f]	0.997	0.813	0.976	0.259	0.901	0.713 (0.006)	0.713-0.714	1000
RF[g]	0.996	0.835	0.976	0.277	0.904	0.727 (0.005)	0.727-0.728	1000
Weighted SVM[h]	0.775	0.819	0.916	0.257	0.685	0.634 (0.005)	0.634-0.634	1000
XGBoost[i]	0.996	0.834	0.976	0.279	0.904	0.728 (0.005)	0.728-0.729	1000

[a]ECE: expected calibration error.
[b]AUC: area under the curve.
[c]CEM: clinical effectiveness metric.
[d]EuroSCORE: European System for Cardiac Operative Risk Evaluation.
[e]LR: logistic regression.
[f]NN: neural network.
[g]RF: random forest.
[h]SVM: support vector machine.
[i]XGBoost: extreme gradient boosting.

Table . The Dunnett test with XGBoost[a] as a control and the rest of the models as comparisons.

Group 1	Group 2 (control)	CEM[b] difference (group 1 – group 2; 95% family-wise CI)	*P* value
EuroSCORE[c] II	XGBoost	−0.1876 (−0.1881 to −0.1870)	$<2\times10^{-16}$[d]
LR[e]	XGBoost	−0.0110 (−0.0116 to −0.0105)	$<2\times10^{-16}$[d]
NN[f]	XGBoost	−0.0148 (−0.0154 to −0.0142)	$<2\times10^{-16}$[d]
RF[g]	XGBoost	−0.0009 (−0.0015 to −0.0003)	.00039[d]
Weighted SVM[h]	XGBoost	−0.0941 (−0.0947 to −0.0935)	$<2\times10^{-16}$[d]

[a]XGBoost: extreme gradient boosting.
[b]CEM: clinical effectiveness metric.
[c]EuroSCORE: European System for Cardiac Operative Risk Evaluation.
[d]*P*<.001.
[e]LR: logistic regression.
[f]NN: neural network.
[g]RF: random forest.
[h]SVM: support vector machine.

Drift Analysis

Overall CEM

Figure 2A shows that XGBoost and RF were candidates for the best overall CEM performance across time. There was minor evidence of LR outperforming NN across time. Seasonal fluctuations were observed. EuroSCORE II performed the worst across time, followed by SVM.

There was strong evidence of a decrease in overall performance across all models (*P*<.0001). Linear regression plots showed that XGBoost had the best starting CEM (intercept: 0.755 vs 0.753 [RF], 0.742 [LR], and 0.741 [NN]), but the rate of performance decrease (slope −0.000720) was less than NN (−0.00083) and greater than RF (−0.000685) and LR (−0.000696; Figure 3A-C and Figure S3 in Multimedia Appendix 1). By March 2019, the overall CEM performance ranking was not changed, with XGBoost performing the best,

followed by RF, LR, and NN. EuroSCORE II (intercept 0.484; slope −0.000847) performed the worst in terms of starting CEM and rate of performance decrease, followed by SVM (intercept 0.658; slope −0.000625; Figure 3D and Figure S4 in Multimedia Appendix 1). Normality and homogeneity assumptions were satisfied for all models' CEM values, as checked by a QQ plot of residuals and scale-location plot (Figure S5 in Multimedia Appendix 1).

Figure 2. (A) Plot of CEM values by model and time. Geometric mean (95% CI) of 1000 bootstraps at each time point is shown. The horizontal line represents the CEM geometric mean of all models. (B) Box plot of difference in models' CEM values across the first 3 months of 2017 and 2019. Kruskal-Wallis results for CEM across the time points are shown. (C) Paired-samples Wilcoxon test (Wilcoxon signed rank test) for the first 3 months of 2019 bootstrap CEM values. *P* values are adjusted using the Bonferroni method. ****$P<.0001$. CEM: clinical effectiveness metric; EuroSCORE: European System for Cardiac Operative Risk Evaluation; ns: not significant; neuronetwork: neural network; SVM: support vector machine; Xgboost: extreme gradient boosting.

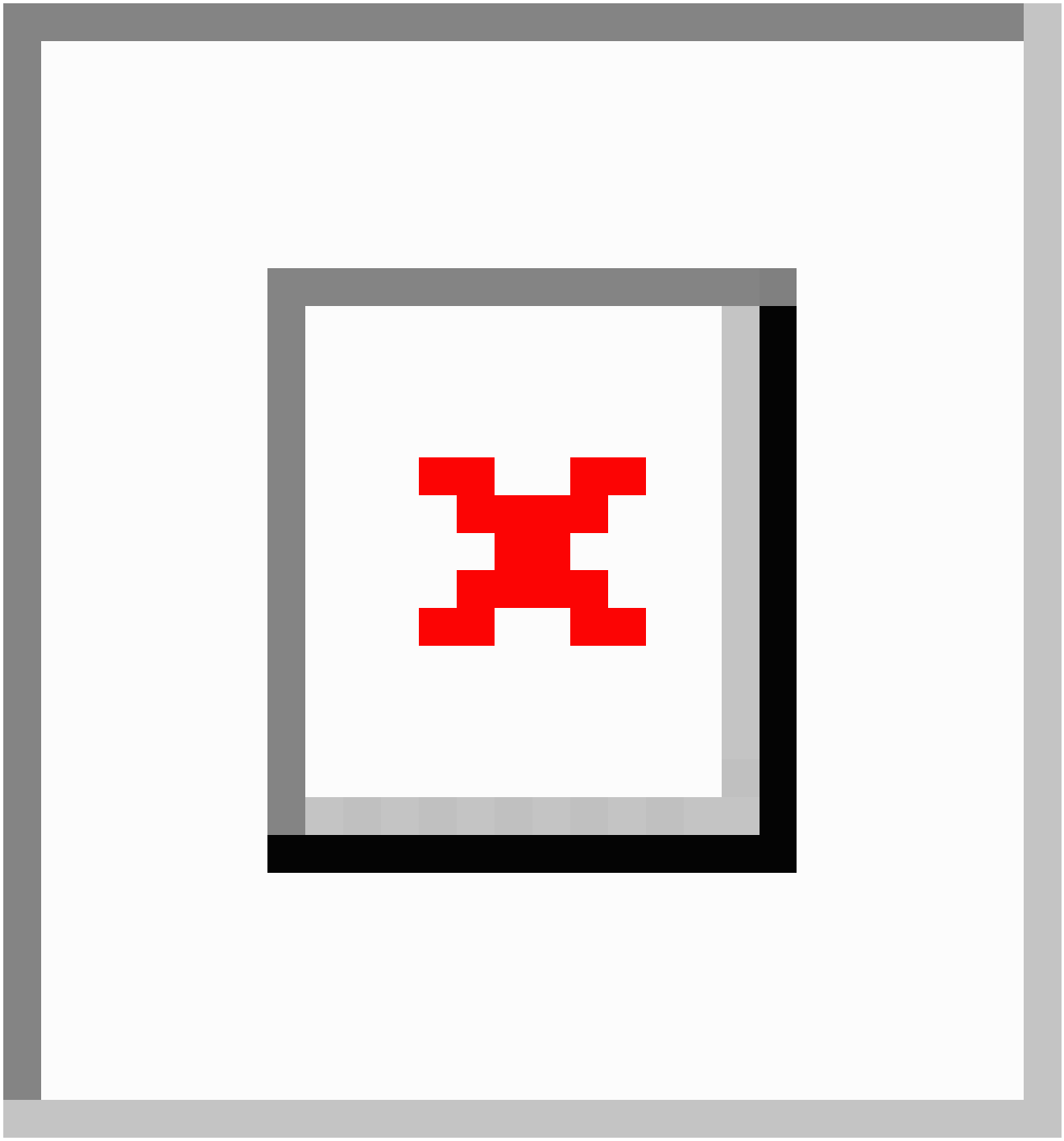
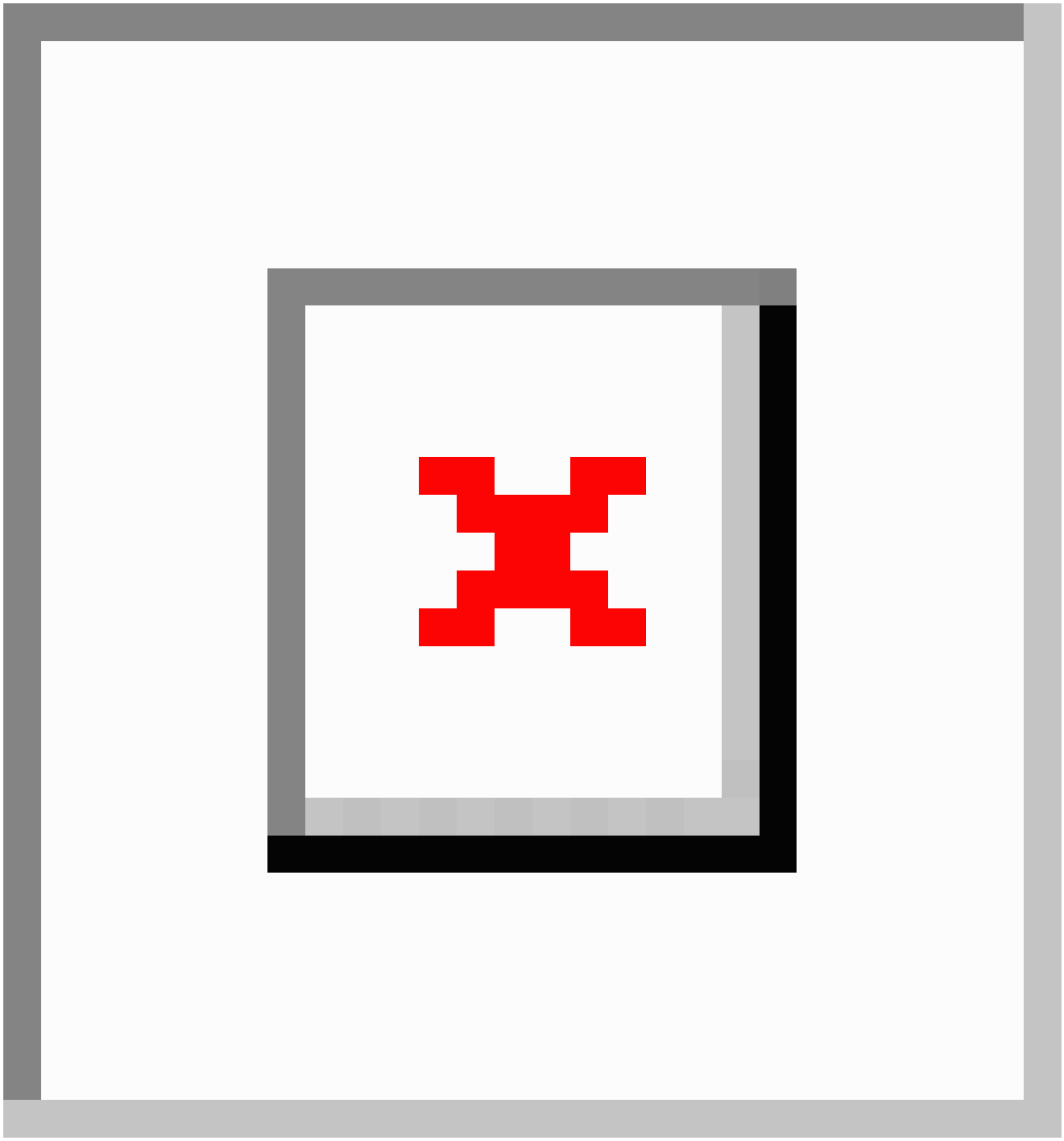


Figure 3. Plots of CEM values by model and time: (A) XGBoost, (B) random forest, (C) logistic regression, and (D) EuroSCORE II. The geometric mean of 1000 bootstraps at each time point is shown. The red dotted line shows linear regression, and the blue line shows generalized additive model fit. Parameters and *P* values for the linear regressions are shown. (E) Discrimination (AUC) performance drift by time. Linear regression lines are plotted for each model, with slope, intercept, and *P* values displayed in the legend. (F) Calibration (adjusted ECE) performance drift by time. Linear regression lines are plotted for each model, with slope, intercept and *P* values displayed in the legend. SVM and EuroSCORE II are removed to enable a clearer separation of models with similar performance. AUC: area under the curve; CEM: clinical effectiveness metric; ECE: expected calibration error; EuroSCORE: European System for Cardiac Operative Risk Evaluation; neuronetwork: neural network; SVM: support vector machine; Xgboost: extreme gradient boosting.



Analysis Within the First 3 Months of 2017

No extreme outliers were found for the models' CEM values in the first 3 months of 2017. The CEM values were nonnormally distributed for all models ($P<.05$; Table S5 in Multimedia Appendix 1). There was strong evidence of a difference across all models ($P<.0001$; Table 3 and Figure S6 in Multimedia Appendix 1). The Dunn test showed strong evidence of XGBoost having the best overall performance (Table S6 in Multimedia Appendix 1; $P<.0001$), followed by RF, NN, and LR (CEM difference to XGBoost: −0.0076, −0.0124, and −0.0138, respectively; $P<.0001$). EuroSCORE II performed the worst, followed by weighted SVM (CEM difference to XGBoost: −0.2739 and −0.0961, respectively; $P<.0001$).

Analysis Within the First 3 Months of 2019

No extreme outliers were found for the models' CEM values in the first 3 months of 2019. The CEM values were nonnormally distributed for 50% (3/6) of models (P<.05). There was strong evidence of a difference across all models (P<.0001; Table S7 in Multimedia Appendix 1 and Figure 2B). The Dunn test showed strong evidence of XGBoost having the best overall performance (Table S8 in Multimedia Appendix 1; P<.05), followed by RF, LR, and NN (CEM difference to XGBoost: −0.0032, −0.0055, and −0.0108, respectively; P<.05). EuroSCORE II performed the worst, followed by weighted SVM (CEM difference to XGBoost: −0.2594 and −0.0856, respectively; P<.0001).

Analysis Between the First 3 Months of 2017 and 2019

No extreme outliers were found for the models' CEM values in the first 3 months of 2017 and 2019. The CEM values were nonnormally distributed for the first 3 months of 2017 and 2019, as assessed by the Kolmogorov-Smirnov test (P<.05). There was strong evidence of an overall difference across the 2 time points (P<.0001; Table S9 and Figure S7 in Multimedia Appendix 1). There was strong evidence of a difference across the 2 time points for each individual model (P<.05; Figure 2C and Table S10 in Multimedia Appendix 1). XGBoost retained the best overall performance across the time points examined. This model showed the largest decrease in CEM performance (median difference 0.0288; P<.0001), followed by NN, RF, and LR (median difference: 0.0272, 0.0244, and 0.0205, respectively; P<.0001). Following a performance decrease from 2017 to 2019, XGBoost still had the best overall performance, with RF being the second best (median CEM: 0.716 and 0.713, respectively). Although NN had a better starting performance than LR, the larger performance drift resulted in NN having a lower overall performance than LR in 2019 (median CEM: 0.705 vs 0.710). Although the performance drift was smaller, LR's CEM performance never exceeded RF's (median CEM: 0.710 vs 0.713). EuroSCORE II showed the least performance drift, followed by weighted SVM (median difference: 0.0142 and 0.0183, respectively; P<.05), but both performed the worst in terms of absolute CEM value.

Analysis of Discrimination, Calibration, and Clinical Effectiveness Drift

Discrimination

AUC

Linear regression plots showed that XGBoost had the best starting AUC (intercept: 0.843 vs 0.839 [RF] and 0.831 [LR, NN, and SVM]), but the rate of performance decrease was greater than RF and EuroSCORE II (slope: −0.000678 vs −0.000381 [RF] and −0.000604 [EuroSCORE II]; Figure 3E). By March 2019, XGBoost's AUC had decreased below RF's, resulting in RF being the best-performing model, followed by XGBoost, SVM, LR, and NN. NN showed the largest rate of AUC decrease, followed by LR and SVM (slope: −0.0014, −0.00093, and −0.000873, respectively). EuroSCORE II performed the worst in terms of AUC across all time points (intercept 0.766). There was strong evidence of a decrease in AUC performance across all models (P<.0001). Normality and homogeneity assumptions were satisfied for all models' AUC values, as checked by a QQ plot of residuals and scale-location plot (Figure S8 in Multimedia Appendix 1).

F_1-score

The best-performing model across all holdout time periods was XGBoost, followed by RF, LR, NN, SVM, and EuroSCORE II. There was strong evidence of a decrease in F_1-score performance across all models (P<.0001). More details can be found in the *Positive Outcome Discrimination* section and Figures S9-10 in Multimedia Appendix 1.

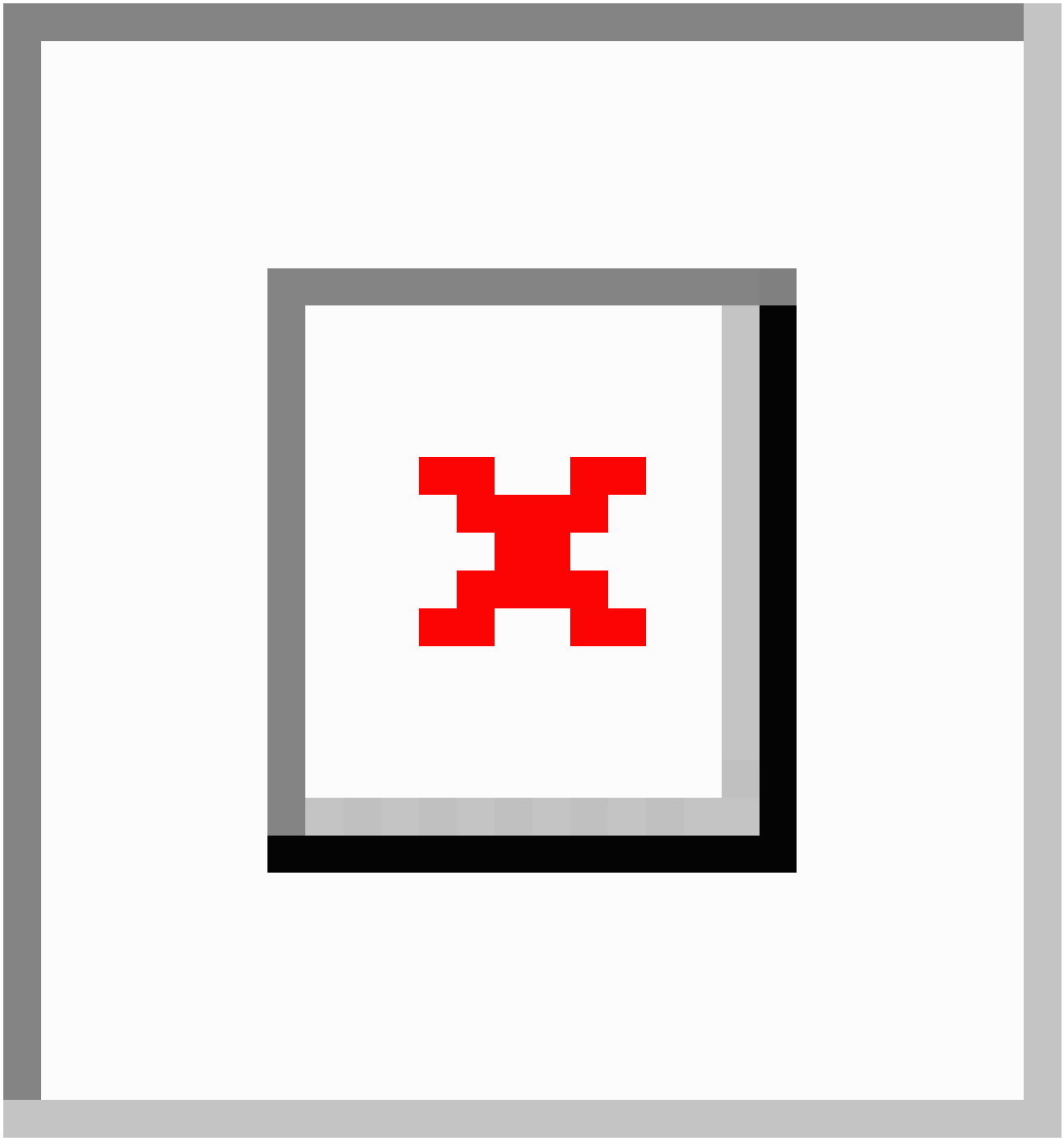
Calibration

Linear regression plots showed that NN has the best starting adjusted ECE (intercept: 0.9907 vs 0.9903 [RF], 0.9902 [XGBoost], and 0.9898 [LR]), but the rate of performance decrease was greater than LR and RF (slope: -5.29×10^{-5} vs -2.93×10^{-6} [LR] and -4.58×10^{-5} [RF]; Figure 3F). By March 2019, NN's adjusted ECE had decreased below LR's, resulting in LR being the best-performing model, followed by NN, RF, and XGBoost. Although SVM and EuroSCORE II had lower rates of adjusted ECE decrease (slope: −0.000251 and −0.000479, respectively), the calibration performance was much lower at all time points compared to the other models (Figure S11 in Multimedia Appendix 1). There was strong evidence of a decrease in adjusted ECE performance across all models (P<.0001), except LR (P>.05). Normality and homogeneity assumptions were satisfied for all models' adjusted ECE values, as checked by a QQ plot of residuals and scale-location plot (Figure S12 in Multimedia Appendix 1).

Clinical Effectiveness

Linear regression plots showed that XGBoost had the best starting net benefit (intercept: 0.9051 vs 0.9043 [RF] and 0.9035 [NN and LR]), but the rate of performance decrease was greater than RF (slope: -5.68×10^{-5} vs -2.5×10^{-6}; Figure 4A), slower than LR (-9.38×10^{-5}), and even slower than NN (−0.000145). By March 2019, XGBoost's net benefit had decreased below RF's, resulting in RF being the best-performing model, followed by XGBoost, LR, and NN. EuroSCORE II showed the largest rate of net benefit decrease and performed the worst across all time points, followed by SVM (intercept: 0.314 and 0.690; slope: −0.000846 and −0.000364, respectively; Figure S13 in Multimedia Appendix 1). There was strong evidence of a decrease in net benefit performance across all models (P<.0001), except RF (P>.05). Normality and homogeneity assumptions were satisfied for all models' net benefit values, as checked by a QQ plot of residuals and scale-location plot (Figure S14 in Multimedia Appendix 1).

Figure 4. (A) Clinical effectiveness (net benefit) performance drift by time. Linear regression lines are plotted for each model, with slope, intercept, and *P* values displayed in the legend. SVM and EuroSCORE II are removed to enable a clearer separation of models with similar performance. (B) SHAP variable importance drift for the holdout set over 27 months (EuroSCORE II and XGBoost). Solid dots show geometric mean values of 5-fold cross-validation. Smoothed locally estimated scatterplot lines are plotted, with green bands showing 95% CIs. (C) SHAP variable importance drift for the holdout set over 27 months for the top 6 most important variables (EuroSCORE II and XGBoost). The trends are unsmoothed. (D) Operative urgency data set drift across time for the holdout set. The percentages of each category are shown for each time point. CCS: Canadian Cardiovascular Society; CPS: critical preoperative state; EuroSCORE: European System for Cardiac Operative Risk Evaluation; ES: EuroSCORE; LV: left ventricle; MI: myocardial infarction; neuronetwork: neural network; NYHA: New York Heart Association; PA: pulmonary artery; PVD: peripheral vascular disease; SHAP: Shapley additive explanations; SVM: support vector machine; Xgboost: extreme gradient boosting.



Accuracy of Prediction Probability

By March 2019, XGBoost was the best model, followed by RF, LR, and NN. EuroSCORE II performed the worst in terms of adjusted Brier score and rate of decrease, followed by SVM. There was strong evidence of a decrease in adjusted Brier score performance across all models ($P<.0001$), except XGBoost and RF. More details can be found in the *Accuracy of Prediction Probability* section and Figures S15-S17 in Multimedia Appendix 1.

Analysis of Variable Importance Drift

SHAP mean absolute magnitude of importance was used to measure variable importance drift for the best temporal and nontemporal model (XGBoost). Smoothed trend lines showed substantial drift in numerous variables, including the most

important variables: age, operative urgency, the weight of intervention, New York Heart Association classification, renal impairment, and previous cardiac surgery (Figure 4B). The sensitivity analysis showed a substantial drift in variable importance across the holdout set for all 6 variables (Figure 4C). When compared with the CEM performance drop from October to December 2017 and from June to July 2018 (Figure 3 generalized additive model line), it could be seen that the CEM decrease was mirrored by decreases in the importance of the top variables, age and operative urgency, at these time periods (Figure 4C). A decrease in CEM performance in the 3 months of 2019 was likely to be at least partly contributed by the sudden rise in the importance of the weight of intervention (Figure 3 and Figure 4B and C).

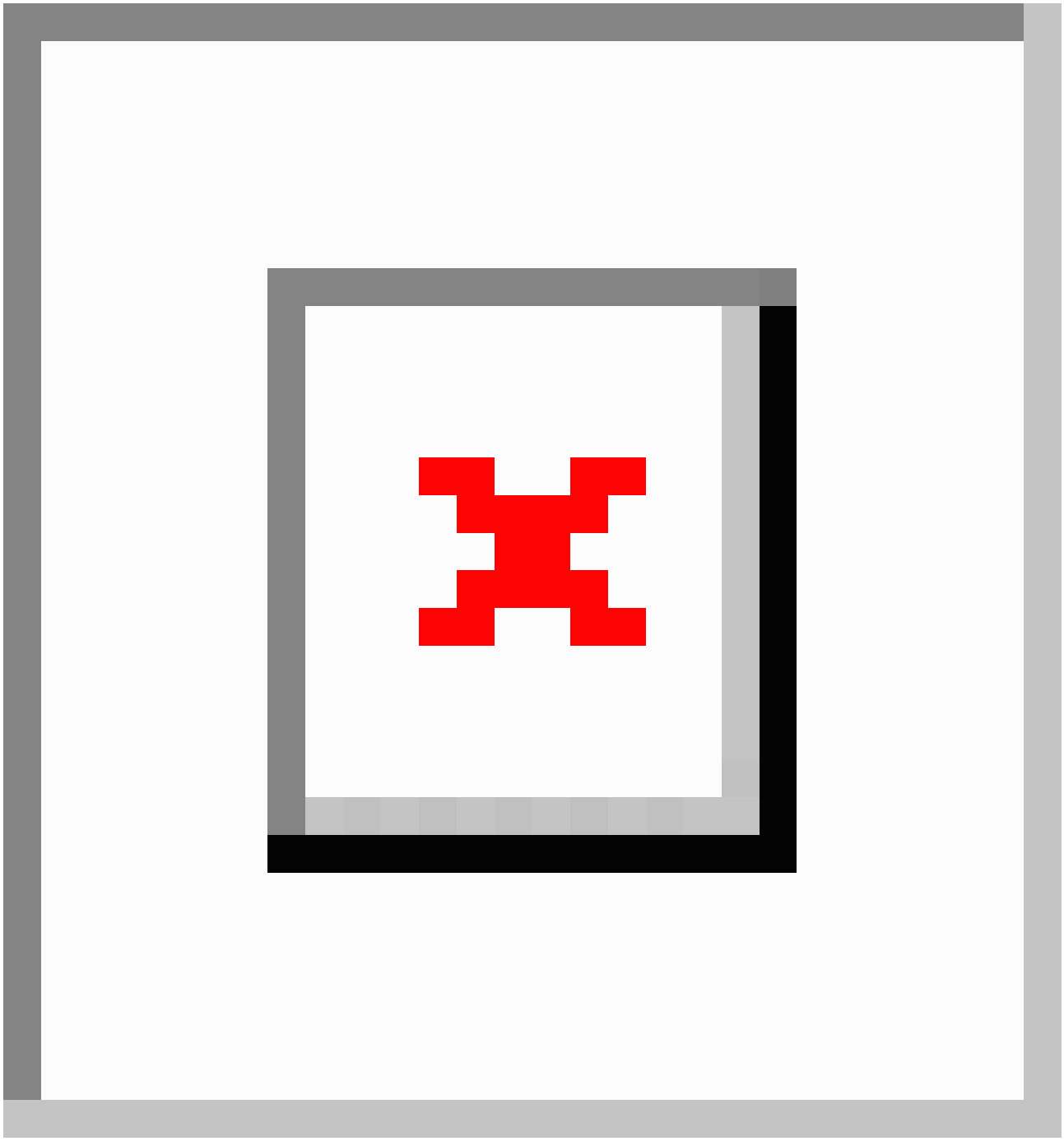
Data Set Drift Across Time

Data set drift was observed throughout the holdout time periods for operative urgency, with sharp drifts observed across all categories from November to December 2017 and from June to July 2018 (Figure 4D). Data set drift was observed across the holdout time periods for the <60 and >60 years patient age groups (Figure S18 in Multimedia Appendix 1), with marked data drifts observed from October to November 2017 and from July to August 2018. Data set drift was observed across the holdout time periods for the weight of intervention (Figure S19 in Multimedia Appendix 1). Sharp data set drifts were observed for the single non-CABG and 3 procedures categories from December 2018 to February 2019.

Net Benefit Projection

To further understand the clinical significance of performance drift over time, Figure 5 illustrates the expected net benefit decrease for the NN and XGBoost models. The blue line depicts the actual net benefit drop for the NN model (as represented by the slope), transitioning to the projected red line after March 2019. The green line represents the actual net benefit drop for the XGBoost model up to March 2019, changing to the projected purple line after March 2019. A clinically significant decrease (from 0.9035 to 0.8808) is shown for NN but not for XGBoost (from 0.9051 to 0.8962).

Figure 5. The actual and projected net benefit drift for the NN and Xgboost models over time. NN: neural network; XGBoost: extreme gradient boosting.



Discussion

Principal Findings

The main finding of the study was that the XGBoost model performed the best, followed by RF, LR, and NN, when all metrics were simultaneously considered, both temporally and nontemporally. Furthermore, EuroSCORE II substantially underperformed against all ML models across all comparisons; this presents an urgent need to understand the drift effects of this score and is not limited to calibration drift. By first combining all metrics and then analyzing the temporal drift of each metric individually, we were able to determine the contribution of individual metrics to the overall performance drift of each model. We found strong evidence that all models showed a decrease in at least 3 of the 5 individual metrics within the CEM. This demonstrated the importance for clinicians and ML governance teams to actively monitor the effects of data set drift (as explained later) on "big data" models that are prepared for or being clinically used to minimize the risk of harm to patients.

"Big data" refers to large and detailed data sets that are suited to ML analyses rather than traditional statistical analyses [45,46], and they are increasingly used in health care. These analyses can inform, personalize, and potentially improve care [45,47,48]. Despite growing interest [49] in ML and health care data linkage initiatives such as the Cardiac Quality Assurance Programme in the United Kingdom [50], there have been limited reports of use within cardiac surgery [51-53], with one of the

main reasons being a lack of understanding by clinicians of the underlining processes [54].

As more countries follow in the steps of the United States to deploy ML to the medical settings [55], it becomes increasingly critical that clinicians and ML governance teams are adequately prepared for situations in which ML systems fail to perform their intended functions [56]. A major factor in ML malfunction is "data set drift," where ML performance declines due to a mismatch between the data on which the model was trained and the new unseen data to which the model is applied [57]. Several factors have been reported to influence data set drift, including changes in technology, demographics, and patient or clinician behavior [56].

In our previous systematic review, we found that despite ML models achieving better discriminatory ability than traditional LR approaches, few cardiac surgery studies assessed calibration, clinical utility, discrimination, and data set drift collectively; these aspects should be assessed to determine the clinical implications of ML [2]. Our previous study [19], although not involving the assessment of XGBoost, had also shown that the calibration drift of LR was less than that of RF, whereas EuroSCORE I, naïve Bayes, and NN performed poorly in terms of calibration. A recent study extending on our work had shown temporal and spatial calibration drift (comparison across regions and hospitals) to be severe across a range of ML models using a national Chinese registry [20]. In accordance with our view, the study highlighted that "future efforts may need to shift more towards enhancing model calibration robustness or recalibration for greater practical value" and that the inclusion of intraoperative variables may be important to enhancing model performance. The STS Adult Cardiac Surgery Database study [21] had shown that the inclusion of intraoperative variables improved both the discrimination and calibration performance of XGBoost and LR models in patients who underwent CABG from the United States. Although calibration drift over time is well documented among EuroSCORE and LR models for hospital mortality, the susceptibility of competing ML modeling methods to data set drift has not been well studied in cardiac surgery [13].

This study heeds the call for additional metrics to address the lack of sensitivity of the most commonly used C-statistic and calibration slope in capturing the advantage of ML models [58]; we demonstrated the use of a consensus score [22,35,59-61] named CEM to take into account numerous metrics that have been found to be beneficial, covering overall accuracy [58], discrimination, calibration, and clinical utility. We wanted to analyze model performance across multiple metrics across time in this study.

This study showed invariance in model ranking for the CEM in both temporal and nontemporal analyses, indicating that there is value for this consensus scoring approach in performance drift evaluation.

This study also addresses the gap in understanding the effect of data set drift on the performance of ML and traditional models over time, which presents a barrier to their clinical application. The shift between XGBoost and RF having the best performance for AUC and net benefit and between NN and LR having the best performance for "adjusted ECE" demonstrates that the comparison of models at a single time point was insufficient to understand the clinical limitations of ML models and that at least 2 time points should be considered.

Our study has also found that although RF showed comparable discrimination (AUC) and clinical utility (net benefit) performance across time, the reason for XGBoost's superior overall temporal performance was in its better overall accuracy (adjusted Brier score) and positive outcome discrimination (F_1-score). F_1-score is often overlooked but is especially important in cardiac surgery data sets, whereby the incidence for the outcome of interest is typically very low and introduces bias in the performance evaluation when AUC is used. We found that RF performed the second best overall. Unlike XGBoost, RF performed better in terms of resistance to drift for AUC and net benefit, suggesting that further work is required to determine whether the synergistic (ensemble) effects across models are beneficial for improving cardiac surgery risk prediction.

Although XGBoost is currently the best temporal and nontemporal model for the NACSA data set, periodic monitoring of performance drift for each yearly revision of this data set should be mandated to determine whether or not performance has been overtaken by RF, and if so, at what point in time this happens [56]. As all models showed strong evidence of a decrease in overall performance from January 2017 to March 2019, further work will be required to develop either better-performing models or models that are less susceptible to performance drift. However, through projecting the net benefit into the year 2030 based on the fitted linear regression, the decreases in the net benefit for XGBoost over time were shown to be clinically insignificant. On the contrary, the NN model showed a clinically significant drop in net benefit.

Although the reported decreases in measures such as CEM and AUC may appear small, such changes are likely to impact the potential use of ML models within clinical scenarios. If such models are to be used clinically for making decisions about the patient, even small changes in these metrics (which have been previously discussed [18] to be important in cardiac surgery ML performance) can have an influence on risk assessment and patient outcomes, necessitating constant model drift monitoring. Prior research has shown that improving model calibration robustness or recalibration is necessary for practical value and that the "the significant decline in performance of previously established models in this study calls for continuing model updates" [20]. It is envisaged that collaboration between physicians and ML scientists is critical. Before mandating model updates, it is critical to establish metric-specific thresholds for acceptable reductions. A consensus approach, extensive experience in this area, or a meta-analysis of current literature may be required for this collaborative decision-making process.

We have demonstrated that by associating relationships between smoothed [62] and unsmoothed trend lines for CEM performance and EuroSCORE II variable importance, it was possible to detect subtle data set drifts that could result in model performance drifts. Our findings of variable importance and data set drift from October to December 2017, from June to July 2018, and from December 2018 to February 2019 are likely

to reflect seasonality changes and mirrored effects of sharp drifts in CEM performance across models. The detection of data set drift was verified by checking for actual drifts in the data set variables. A noncardiac surgery study used actual data set drift to check for variable importance–detected data set drift [13]. However, drift in the actual data set was only analyzed across 2 data points [13], without consideration for smoothed and unsmoothed relationships across performance, variable importance, and actual variable incidence. This study provides the foundations for which further work analyzing ML performance drift are recommended, to analyze relationships between drifts in a consensus score such as CEM and in variable importance, followed by the confirmation of any detected drifts using actual data set trends (data set drift).

Limitations and Future Studies

Although statistical rigor has been applied to determine whether performance drift is a barrier to clinical risk modeling and decision-making, further work could be done to apply more statistically sensitive approaches for comparing the interactions of trends in data set drift, performance drift, and variable importance drift. As NACSA patient identifiers and the Hospital Episode Statistics data set were not available for linkage, it was not possible to determine whether there were any same patient individuals in both the training and validation set and holdout set, where they had multiple surgeries. Clinical judgment suggests that the proportion of multiple surgeries would be very low. Nonetheless, future work should consider the collection of such information to minimize any potential bias. Our previous work using CEM and constituent metrics to study random effects ML had also shown that hospital-related systematic variations may be better adjusted for by including hospital location variables as part of the input covariates rather than specifically using mixed effects ML models [17]. Future work may consider the incorporation of such systematic variation adjustments when studying drift effects to further investigate the optimal approach for modeling drift across individual hospitals. Although CEM is a consensus score that enhances the clinical evaluation of complex relationships across different aspects of model performance, compressing the net benefit measure into a single value would mean that further DCA may be required if individual-specific, threshold-based decisions were to be fully considered. Future studies should also delve deeper into the relationships of the studied drift types with concept drift in cardiac surgery risk prediction.

Conclusion

This study found that performance drift of ML and EuroSCORE II over time could be explained through data set drift patterns in cardiac surgery risk prediction. It was also found that variable importance drift could help to explain performance drift and support the detection of data set drift in the assessed models. The strong evidence of all models showing a decrease in at least 3 of the 5 individual metrics within CEM demonstrates the potential need to update the models over time, but future work are required to determine suitable thresholds for mandating an update. Future work will be required to determine the interplay between XGBoost and RF, which have demonstrated less drift over time, and whether combining these models through additional ensemble modeling could take advantage of their respective performance advantages.

Acknowledgments

This work was supported by a grant from the British Heart Foundation–Turing Institute, the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at University Hospitals Bristol, Weston NHS Foundation Trust, and the University of Bristol.

Data Availability

All data used in this study are from the National Adult Cardiac Surgery Audit (NACSA) data set. These data may be requested from the Healthcare Quality Improvement Partnership (HQIP) [63]. Code for deriving training, update, and holdout data sets is available on GitHub upon reasonable request to the corresponding author, and the authors can provide confirmatory deidentified record IDs for each set upon reasonable request.

Authors' Contributions

TD, SS, AD, DF, JC, BZ, PN, UB, AJ, and GA contributed to the experimental design. TD and SS acquired the data. TD and SS performed the data preprocessing. TD wrote the source code to perform the experiments and is accountable for all aspects of the work. TD, SS, AD, DF, JC, BZ, PN, AJ, and GA analyzed the results. TD wrote the first version of the paper. All authors revised the paper and approved the submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data set split, model specification, drift analysis, and other analyses.
[DOCX File, 144155 KB - xmed_v5i1e45973_app1.docx]

References

1. Ong CS, Reinertsen E, Sun H, et al. Prediction of operative mortality for patients undergoing cardiac surgical procedures without established risk scores. J Thorac Cardiovasc Surg 2023 Apr;165(4):1449-1459. [doi: 10.1016/j.jtcvs.2021.09.010] [Medline: 34607725]
2. Benedetto U, Dimagli A, Sinha S, et al. Machine learning improves mortality risk prediction after cardiac surgery: systematic review and meta-analysis. J Thorac Cardiovasc Surg 2022 Jun;163(6):2075-2087. [doi: 10.1016/j.jtcvs.2020.07.105] [Medline: 32900480]
3. Kieser TM, Rose MS, Head SJ. Comparison of logistic EuroSCORE and EuroSCORE II in predicting operative mortality of 1125 total arterial operations. Eur J Cardiothorac Surg 2016 Sep;50(3):509-518. [doi: 10.1093/ejcts/ezw072] [Medline: 27005979]
4. Poullis M, Pullan M, Chalmers J, Mediratta N. The validity of the original EuroSCORE and EuroSCORE II in patients over the age of seventy. Interact Cardiovasc Thorac Surg 2015 Feb;20(2):172-177. [doi: 10.1093/icvts/ivu345] [Medline: 25348730]
5. Zhang GX, Wang C, Wang L, et al. Validation of EuroSCORE II in Chinese patients undergoing heart valve surgery. Heart Lung Circ 2013 Aug;22(8):606-611. [doi: 10.1016/j.hlc.2012.12.012] [Medline: 23375874]
6. Silaschi M, Conradi L, Seiffert M, et al. Predicting risk in transcatheter aortic valve implantation: comparative analysis of EuroSCORE II and established risk stratification tools. Thorac Cardiovasc Surg 2015 Sep;63(6):472-478. [doi: 10.1055/s-0034-1389107] [Medline: 25191764]
7. Carnero-Alcázar M, Silva Guisasola JA, Reguillo Lacruz FJ, et al. Validation of EuroSCORE II on a single-centre 3800 patient cohort. Interact Cardiovasc Thorac Surg 2013 Mar;16(3):293-300. [doi: 10.1093/icvts/ivs480] [Medline: 23178391]
8. Arangalage D, Cimadevilla C, Alkhoder S, et al. Agreement between the new EuroSCORE II, the logistic EuroSCORE and the Society of Thoracic Surgeons score: implications for transcatheter aortic valve implantation. Arch Cardiovasc Dis 2014;107(6-7):353-360. [doi: 10.1016/j.acvd.2014.05.002] [Medline: 24996564]
9. Atashi A, Amini S, Tashnizi MA, et al. External validation of European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) for risk prioritization in an Iranian population. Braz J Cardiovasc Surg 2018;33(1):40-46. [doi: 10.21470/1678-9741-2017-0030] [Medline: 29617500]
10. Provenchère S, Chevalier A, Ghodbane W, et al. Is the EuroSCORE II reliable to estimate operative mortality among octogenarians? PLoS One 2017 Nov 16;12(11):e0187056. [doi: 10.1371/journal.pone.0187056] [Medline: 29145434]
11. Nilsson J, Ohlsson M, Thulin L, Höglund P, Nashef SAM, Brandt J. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. J Thorac Cardiovasc Surg 2006 Jul;132(1):12-19. [doi: 10.1016/j.jtcvs.2005.12.055] [Medline: 16798296]
12. Kurlansky P. Commentary: the risk of risk models. J Thorac Cardiovasc Surg 2020 Jul;160(1):181-182. [doi: 10.1016/j.jtcvs.2019.12.063] [Medline: 32007255]
13. Duckworth C, Chmiel FP, Burns DK, et al. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. Sci Rep 2021 Nov 26;11(1):23017. [doi: 10.1038/s41598-021-02481-y] [Medline: 34837021]
14. Dong T, Sinha S, Zhai B, et al. Cardiac surgery risk prediction using ensemble machine learning to incorporate legacy risk scores: a benchmarking study. Digit Health 2023 Jul 20;9:20552076231187605. [doi: 10.1177/20552076231187605] [Medline: 37492033]
15. Hickey GL, Blackstone EH. External model validation of binary clinical risk prediction models in cardiovascular and thoracic surgery. J Thorac Cardiovasc Surg 2016 Aug;152(2):351-355. [doi: 10.1016/j.jtcvs.2016.04.023] [Medline: 27215928]
16. Kang X. The effect of color on short-term memory in information visualization. In: VINCI '16: Proceedings of the 9th International Symposium on Visual Information Communication and Interaction: Association for Computing Machinery; 2016:144-145. [doi: 10.1145/2968220.2968237]
17. Dong T, Sinha S, Fudulu DP, et al. Random effects adjustment in machine learning models for cardiac surgery risk prediction: a benchmarking study. medRxiv. Preprint posted online on Jun 12, 2023. [doi: 10.1101/2023.06.08.23291129]
18. Sinha S, Dong T, Dimagli A, et al. Comparison of machine learning techniques in prediction of mortality following cardiac surgery: analysis of over 220 000 patients from a large national database. Eur J Cardiothorac Surg 2023 Jun 1;63(6):ezad183. [doi: 10.1093/ejcts/ezad183] [Medline: 37154705]
19. Benedetto U, Sinha S, Lyon M, et al. Can machine learning improve mortality prediction following cardiac surgery? Eur J Cardiothorac Surg 2020 Dec 1;58(6):1130-1136. [doi: 10.1093/ejcts/ezaa229] [Medline: 32810233]
20. Zeng J, Zhang D, Lin S, et al. Comparative analysis of machine learning vs. traditional modeling approaches for predicting in-hospital mortality after cardiac surgery: temporal and spatial external validation based on a nationwide cardiac surgery registry. Eur Heart J Qual Care Clin Outcomes 2024 Mar 1;10(2):121-131. [doi: 10.1093/ehjqcco/qcad028] [Medline: 37218710]
21. Mori M, Durant TJS, Huang C, et al. Toward dynamic risk prediction of outcomes after coronary artery bypass graft: improving risk prediction with intraoperative events using gradient boosting. Circ Cardiovasc Qual Outcomes 2021 Jun;14(6):e007363. [doi: 10.1161/CIRCOUTCOMES.120.007363] [Medline: 34078100]

22. Dong T, Benedetto U, Sinha S, et al. Deep recurrent reinforced learning model to compare the efficacy of targeted local versus national measures on the spread of COVID-19 in the UK. BMJ Open 2022 Feb 21;12(2):e048279. [doi: 10.1136/bmjopen-2020-048279] [Medline: 35190408]
23. Kamaleswaran R, Mahajan R, Akbilgic O. A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length. Physiol Meas 2018 Mar 27;39(3):035006. [doi: 10.1088/1361-6579/aaaa9d] [Medline: 29369044]
24. Zou F, Shen L, Jie Z, Zhang W, Liu W. A sufficient condition for convergences of Adam and RMSProp. Presented at: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Jun 15 to 20, 2019:; Long Beach, CA p. 11119-11127. [doi: 10.1109/CVPR.2019.01138]
25. Chakraborty D, Awolusi I, Gutierrez L. An explainable machine learning model to predict and elucidate the compressive behavior of high-performance concrete. Results Eng 2021 Sep;11:100245. [doi: 10.1016/j.rineng.2021.100245]
26. Hickey GL, Grant SW, Cosgriff R, et al. Clinical registries: governance, management, analysis and applications. Eur J Cardiothorac Surg 2013 Oct;44(4):605-614. [doi: 10.1093/ejcts/ezt018] [Medline: 23371972]
27. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci 2017 Oct 6;9:329. [doi: 10.3389/fnagi.2017.00329] [Medline: 29056906]
28. Prabhakararao E, Dandapat S. A weighted SVM based approach for automatic detection of posterior myocardial infarction using VCG signals. Presented at: 2019 National Conference on Communications (NCC); Feb 20 to 23, 2019:; Bangalore, India p. 1-6. [doi: 10.1109/NCC.2019.8732238]
29. Rajliwall NS, Davey R, Chetty G. Cardiovascular risk prediction based on XGBoost. Presented at: 2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE); Dec 10 to 12, 2018:; Nadi, Fiji p. 246-252. [doi: 10.1109/APWConCSE.2018.00047]
30. Kumar NK, Sindhu GS, Prashanthi DK, Sulthana AS. Analysis and prediction of cardio vascular disease using machine learning classifiers. Presented at: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS); Mar 6 to 7, 2020:; Coimbatore, India p. 15-21. [doi: 10.1109/ICACCS48705.2020.9074183]
31. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. JAMA Netw Open 2020 Jan 3;3(1):e1919396. [doi: 10.1001/jamanetworkopen.2019.19396] [Medline: 31951272]
32. Allyn J, Allou N, Augustin P, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. PLoS One 2017 Jan 6;12(1):e0169772. [doi: 10.1371/journal.pone.0169772] [Medline: 28060903]
33. Mehrtash A, Wells WM, Tempany CM, Abolmaesumi P, Kapur T. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. IEEE Trans Med Imaging 2020 Dec;39(12):3868-3878. [doi: 10.1109/TMI.2020.3006437] [Medline: 32746129]
34. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138. [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
35. Devaraj J, Madurai Elavarasan R, Pugazhendhi R, et al. Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant? Results Phys 2021 Feb;21:103817. [doi: 10.1016/j.rinp.2021.103817] [Medline: 33462560]
36. Rohrbach L, Huebner P. Application of multi-criteria decision making in bioink selection. Presented at: 2021 Systems and Information Engineering Design Symposium (SIEDS); Apr 29 to 30, 2021:; Charlottesville, VA p. 1-6. [doi: 10.1109/SIEDS52267.2021.9483762]
37. Armstrong JS, Collopy F. Error measures for generalizing about forecasting methods: empirical comparisons. Int J Forecast 1992 Jun;8(1):69-80. [doi: 10.1016/0169-2070(92)90008-W]
38. Kacalak W, Lipiński D, Różański R, Królczyk GM. Assessment of the classification ability of parameters characterizing surface topography formed in manufacturing and operation processes. Measurement 2021 Jan;170:108715. [doi: 10.1016/j.measurement.2020.108715]
39. Krejčí J, Stoklasa J. Aggregation in the analytic hierarchy process: why weighted geometric mean should be used instead of weighted arithmetic mean. Expert Syst Appl 2018 Dec 30;114:97-106. [doi: 10.1016/j.eswa.2018.06.060]
40. González-Estrada E, Cosmes W. Shapiro–Wilk test for skew normal distributions based on data transformations. J Stat Comput Simul 2019 Aug 27;89(17):3258-3272. [doi: 10.1080/00949655.2019.1658763]
41. Guidance for data quality assessment. United States Environmental Protection Agency. URL: https://www.epa.gov/quality/guidance-data-quality-assessment [accessed 2022-02-10]
42. McLeod AI. Improved spread-location visualization. J Comput Graph Stat 1999;8(1):135-141. [doi: 10.1080/10618600.1999.10474806]
43. Barda N, Riesel D, Akriv A, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nat Commun 2020 Sep 7;11(1):4439. [doi: 10.1038/s41467-020-18297-9] [Medline: 32895375]
44. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, von Luxburg U, Bengio S, et al, editors. Advances in Neural Information Processing Systems 30 (NIPS 2017): Curran Associates, Inc; 2017:1-10 URL:

https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html [accessed 2024-05-23]
45. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014 Feb 7;2:3. [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
46. Silverio A, Cavallo P, de Rosa R, Galasso G. Big health data and cardiovascular diseases: a challenge for research, an opportunity for clinical care. Front Med (Lausanne) 2019 Feb 25;6:36. [doi: 10.3389/fmed.2019.00036] [Medline: 30873409]
47. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. Heredity (Edinb) 2020 Apr;124(4):525-534. [doi: 10.1038/s41437-020-0303-2] [Medline: 32139886]
48. Pencina MJ, Goldstein BA, D'Agostino RB. Prediction models — development, evaluation, and clinical application. N Engl J Med 2020 Apr 23;382(17):1583-1586. [doi: 10.1056/NEJMp2000589] [Medline: 32320568]
49. Ruiz VM, Goldsmith MP, Shi L, et al. Early prediction of clinical deterioration using data-driven machine-learning modeling of electronic health records. J Thorac Cardiovasc Surg 2022 Jul;164(1):211-222.e3. [doi: 10.1016/j.jtcvs.2021.10.060] [Medline: 34949457]
50. Kendall S, Shah R, Moorjani N, et al. Adult cardiac surgery quality improvement and quality assurance: proposal to improve the audit. 2021 to 2024 and beyond. Society for Cardiothoracic Surgery in Great Britain and Ireland. 2020 Oct 6. URL: https://scts.org/_userfiles/pages/files/adult%20cardiac/scts_adult_cardiac_surgery_accreditation_proposal_202124.pdf [accessed 2024-05-23]
51. Hernandez-Suarez DF, Kim Y, Villablanca P, et al. Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. JACC Cardiovasc Interv 2019 Jul 22;12(14):1328-1338. [doi: 10.1016/j.jcin.2019.06.013] [Medline: 31320027]
52. Wojnarski CM, Roselli EE, Idrees JJ, et al. Machine-learning phenotypic classification of bicuspid aortopathy. J Thorac Cardiovasc Surg 2018 Feb;155(2):461-469.e4. [doi: 10.1016/j.jtcvs.2017.08.123] [Medline: 29042101]
53. Chen Z, Li J, Sun Y, et al. A novel predictive model for poor in-hospital outcomes in patients with acute kidney injury after cardiac surgery. J Thorac Cardiovasc Surg 2023 Mar;165(3):1180-1191. [doi: 10.1016/j.jtcvs.2021.04.085] [Medline: 34112503]
54. Domaratzki M, Kidane B. Deus ex machina? demystifying rather than deifying machine learning. J Thorac Cardiovasc Surg 2022 Mar;163(3):1131-1137.e4. [doi: 10.1016/j.jtcvs.2021.02.095] [Medline: 33840471]
55. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 4;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
56. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med 2021 Jul 15;385(3):283-286. [doi: 10.1056/NEJMc2104626] [Medline: 34260843]
57. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. Biostatistics 2020 Apr 1;21(2):345-352. [doi: 10.1093/biostatistics/kxz041] [Medline: 31742354]
58. Huang C, Li SX, Caraballo C, et al. Performance metrics for the comparative analysis of clinical risk prediction models employing machine learning. Circ Cardiovasc Qual Outcomes 2021 Oct;14(10):e007526. [doi: 10.1161/CIRCOUTCOMES.120.007526] [Medline: 34601947]
59. Ericksen SS, Wu H, Zhang H, et al. Machine learning consensus scoring improves performance across targets in structure-based virtual screening. J Chem Inf Model 2017 Jul 24;57(7):1579-1590. [doi: 10.1021/acs.jcim.7b00153] [Medline: 28654262]
60. Hornik K, Meyer D. Deriving consensus rankings from benchmarking experiments. In: Decker R, Lenz HJ, editors. Advances in Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization: Springer; 2007:163-170. [doi: 10.1007/978-3-540-70981-7_19]
61. Hu J, Peng Y, Lin Q, Liu H, Zhou Q. An ensemble weighted average conservative multi-fidelity surrogate modeling method for engineering optimization. Eng Comput 2022 Jun;38(3):2221-2244. [doi: 10.1007/s00366-020-01203-8]
62. Fudulu DP, Dimagli A, Sinha S, et al. Weekday and outcomes of elective cardiac surgery in the UK: a large retrospective database analysis. Eur J Cardiothorac Surg 2022 May 27;61(6):1381-1388. [doi: 10.1093/ejcts/ezac038] [Medline: 35092280]
63. Data access at HQIP. Healthcare Quality Improvement Partnership. URL: https://www.hqip.org.uk/national-programmes/accessing-ncapop-data/#.Ys6gN-zMLdp [accessed 2024-05-23]

Abbreviations

AUC: area under the curve
CABG: coronary artery bypass grafting
CEM: clinical effectiveness metric
CONSORT: Consolidated Standards of Reporting Trials
DCA: decision curve analysis
ECE: expected calibration error
EuroSCORE: European System for Cardiac Operative Risk Evaluation
LR: logistic regression

ML: machine learning
NACSA: National Adult Cardiac Surgery Audit
NN: neural network
RF: random forest
ROC: receiver operating characteristic
SHAP: Shapley additive explanations
SinoSCORE II: Sino (Chinese) System for Coronary Artery Bypass Grafting Operative Risk Evaluation II
STS: Society of Thoracic Surgeons
SVM: support vector machine
XGBoost: extreme gradient boosting

Edited by E Meinert; submitted 08.06.23; peer-reviewed by Anonymous, J Zeng; revised version received 27.02.24; accepted 29.04.24; published 12.06.24.

Please cite as:
Dong T, Sinha S, Zhai B, Fudulu D, Chan J, Narayan P, Judge A, Caputo M, Dimagli A, Benedetto U, Angelini GD
Performance Drift in Machine Learning Models for Cardiac Surgery Risk Prediction: Retrospective Analysis
JMIRx Med 2024;5:e45973
URL: https://xmed.jmir.org/2024/1/e45973
doi:10.2196/45973

Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis

Tigran Makunts[1], PharmD; Haroutyun Joulfayan[2]; Ruben Abagyan[1], PhD

1

2

Corresponding Author:
Ruben Abagyan, PhD

Related Articles:

Companion article: https://preprints.jmir.org/preprint/55976

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.19.23298750v1

Companion article: https://med.jmirx.org/2024/1/e59120

Companion article: https://med.jmirx.org/2024/1/e58273

Abstract

Background: Glucagon-like peptide-1 (GLP-1) receptor agonists (RAs) are one of the most commonly used drugs for type 2 diabetes mellitus. Clinical guidelines recommend GLP-1 RAs as an adjunct to diabetes therapy in patients with chronic kidney disease, presence or risk of atherosclerotic cardiovascular disease, and obesity. The weight loss observed in clinical trials has been explored further in healthy individuals, putting GLP-1 RAs on track to be the next weight loss treatment.

Objective: Although the adverse event profile is relatively safe, most GLP-1 RAs come with a labeled boxed warning for the risk of thyroid cancers, based on animal models and some postmarketing case reports in humans. Considering the increasing popularity of this drug class and its expansion into a new popular indication, a further review of the most recent postmarketing safety data was warranted to quantify thyroid hyperplasia and neoplasm instances.

Methods: GLP-1 RA patient reports from the US Food and Drug Administration (FDA) Adverse Event Reporting System database were analyzed using reporting odds ratios and 95% CIs.

Results: In this study, we analyzed over 18 million reports from the US FDA Adverse Event Reporting System and provided evidence of significantly increased propensity for thyroid hyperplasias and neoplasms in patients taking GLP-1 RA monotherapy when compared to patients taking sodium-glucose cotransporter-2 (SGLT-2) inhibitor monotherapy.

Conclusions: GLP-1 RAs, regardless of indication, are associated with an over 10-fold increase in thyroid neoplasm and hyperplasia adverse event reporting when compared to SGLT-2 inhibitors.

(JMIRx Med 2024;5:e55976) doi:10.2196/55976

KEYWORDS

GLP-1; FDA; averse event reporting; cancer; oncology; neoplasm; drugs; pharmacy; pharmacology; pharmaceutics; medication; medications; glucagon-like peptide-1; Food and Drug Administration; weight loss; diabetes; obesity; thyroid hyperplasia; FAERS; FDA Adverse Event Reporting System

Introduction

Glucagon-like peptide-1 (GLP-1) receptor agonists (RAs) have gained increased popularity due to the improved safety and efficacy profiles observed in clinical trials. This class of drugs includes liraglutide, semaglutide, exenatide, and dulaglutide [1-5]. GLP-1 RAs are indicated as a therapeutic adjunct to diet and exercise, to improve glycemic control in patients with type 2 diabetes mellitus (T2DM). The American Association of

Clinical Endocrinology and the American Diabetes Association recommend GLP-1 RAs for patients with T2DM who have chronic kidney disease, atherosclerotic cardiovascular disease risk, or obesity [6,7]. GLP-1 RAs have also been used in a wide range of cardiometabolic conditions, including but not limited to nonalcoholic fatty liver disease and nonalcoholic steatohepatitis [8].

GLP-1 RAs were recommended for patients with both T2DM and obesity because of the weight loss observed during the clinical trials [2-5,9]. The effect was attributed to decreasing gastric emptying, peristalsis, appetite, and glucose absorption [10,11]. This effect was further explored in studies of populations without T2DM [12-14], suggesting the expansion of GLP-1 RA use.

Common adverse events (AEs) associated with GLP-1 RAs observed during the clinical trials include nausea, hypoglycemia, vomiting, diarrhea, feeling jittery, dizziness, headache, and dyspepsia. Of a greater concern are the labeled boxed warnings of liraglutide, semaglutide, and dulaglutide, marking these as contraindicated in patients with a family history of medullary thyroid carcinoma. These warnings were based on nonhuman data about the development of thyroid C-cell tumors in rats and mice receiving clinically relevant doses of GLP-1 RAs [15-17]. The thyroid cancer association in humans has been studied and observed as well in retrospective studies [18-20]. However, there is conflicting evidence from a meta-analysis of human randomized controlled trials, which refutes this association [21].

The increased popularity of GLP-1 RAs and the unsettled association of thyroid hyperplasias and neoplasms prompted further investigation into the most recent US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) data sets. In this study, we evaluated thyroid hyperplasia and neoplasm–related AEs that are reported as being associated with GLP-1 RA monotherapy when compared to sodium-glucose cotransporter-2 (SGLT-2) inhibitors monotherapy. SGLT-2 inhibitors, a class of drugs that work by blocking renal glucose reabsorption, were selected as the control cohort due to their comparable indication in diabetes treatment guidelines. We analyzed GLP-1 RAs individually and as a class.

Methods

Ethical Considerations

Ethical review and approval were not required for the study on human participants in accordance with local legislation and institutional requirements. Written informed consent from the participants' legal guardian or next of kin was not required to participate in this study in accordance with national legislation and institutional requirements. The data used for the analysis had been deidentified and made public by the US FDA.

FAERS Data Sets

The FAERS is a repository of AE cases sent to the FDA through MedWatch (form 3500/3500a) [22,23]. The cases include AEs submitted voluntarily by health care professionals, individuals, and legal representatives and spontaneous mandatory reports by manufacturers and sponsors. At the time of the analysis, the FAERS contained 18,274,795 reports from January 2004 to September 2022.

The data sets have been deidentified and made available on the web [24].

Data Preparation, Cohort Selection, and Outcome Measure

The FAERS quarterly data sets were initially downloaded in text format. Due to the variability of data structure between quarters and the paucity of some of the variables, the cases were standardized to fit a uniform structure [25-27].

Out of a total of 18,274,795 reports, AE cases where the report was submitted by a health care professional (pharmacists, physicians, nurses, or other health care professionals) were selected as the initial data set (n=6,360,489). This allowed for minimizing reporting bias and adding to the clinical relevance of the studied cases. Further, reports where GLP-1 RAs were the only drug reported (further referred to as *monotherapy*) were selected as the GLP-1 RA cohort (n=17,653) to avoid potential confounding effects from concomitant medications. The cohort was further split into individual GLP-1 RA subcohorts: semaglutide (n=3230), dulaglutide (n=3768), exenatide (n=4493), and liraglutide (n=6162). SGLT-2 inhibitor monotherapy reports (canagliflozin, dapagliflozin, and empagliflozin; n=14,102) were selected as a control cohort. SGLT-2 inhibitors were selected as a control due to the comparable recommendation by diabetes management guidelines. Metformin monotherapy was initially selected as the control cohort to match the population of the cohort of interest (n=8536); however, only a single case of interest (Preferred Term [PT] code: thyroid cancer) was reported. The undetectable baseline of this AE made it impossible to use metformin as a control (Table 1). All the thyroid-related AE PT codes based on standard MedDRA queries and FDA Medical queries were used in the case selection process. Disproportionality analysis using reporting odds ratios (RORs) and 95% CIs was used to determine the statistical significance of the results. Of the 31,755 included reports, only those from health care professionals (pharmacists: n=3620, 11.4%; physicians: n=20,133, 63.4%; and other health care professionals: n=8002, 25.2%) were included in the analysis. The reports were primarily from the United States (United States: n=26,452, 83.3%; Japan: n=603, 1.9%; United Kingdom: n=540, 1.7%; France: n=476, 1.5%; Canada: n=349, 1.1%; Australia: n=349, 1.1%; and other counties: n=2986, 9.4%, <1% in each country). The reports were from the following years: 2010 (n=2128, 6.7%), 2011 (n=3430, 10.8%), 2012 (n=2096, 6.6%), 2013 (n=1143, 3.6%), 2014 (n=1080, 3.4%), 2015 (n=1969, 6.2%), 2016 (n=1429, 4.5%), 2017 (n=1651, 5.2%), 2018 (n=2159, 6.8%), 2019 (n=2191, 6.9%), 2020 (n=4096, 12.9%), 2021 (n=4954, 15.6%), and 2022 (n=3430, 10.8%; only the first 3 quarters at the time of the analysis).

Table . Thyroid hyperplasia or neoplasm–related AE[a] reports in GLP-1[b] RA[c], metformin, and SGLT-2[d] inhibitor FAERS[e] reports.

AE Preferred Term	AE reports, n (%)					
	Semaglutide (n=3230)	Dulaglutide (n=3768)	Exenatide (n=4493)	Liraglutide (n=6162)	Metformin (n=8536)	SGLT-2 inhibitors (canagliflozin, empagliflozin, and dapagliflozin; control[f]; n=14,102)
Thyroid mass	19 (0.59)	9 (0.24)	5 (0.11)	25 (0.41)	0 (0)	1 (0.01)
Medullary thyroid cancer	8 (0.25)	5 (0.13)	1 (0.02)	7 (0.11)	0 (0)	2 (0.01)
Thyroid cancer	3 (0.09)	10 (0.27)	4 (0.09)	30 (0.49)	1 (0.01)	3 (0.02)
Papillary thyroid cancer	2 (0.06)	3 (0.08)	2 (0.04)	20 (0.32)	0 (0)	0 (0)
Benign neoplasm of thyroid gland	2 (0.06)	2 (0.05)	2 (0.04)	1 (0.02)	0 (0)	1 (0.01)
Thyroid neoplasm	0 (0)	2 (0.05)	3 (0.07)	13 (0.21)	0 (0)	1 (0.01)
Thyroid cyst	0 (0)	2 (0.05)	0 (0)	6 (0.10)	0 (0)	0 (0)
Follicular thyroid cancer	0 (0)	0 (0)	1 (0.02)	1 (0.02)	0 (0)	0 (0)
Thyroid adenoma	0 (0)	0 (0)	0 (0)	4 (0.06)	0 (0)	0 (0)
Thyroid C-cell hyperplasia	0 (0)	0 (0)	0 (0)	1 (0.02)	0 (0)	0 (0)
Thyroid cancer metastatic	0 (0)	0 (0)	0 (0)	2 (0.03)	0 (0)	0 (0)
Unique individual thyroid hyperplasia or neoplasm–related AE reports[g]	*33 (1.02)*	*33 (0.88)*	*17 (0.38)*	*108 (1.75)*	*1 (0.01)*	*7 (0.05)*

[a]AE: adverse event.

[b]GLP-1: glucagon-like peptide-1.

[c]RA: receptor agonist.

[d]SGLT-2: sodium-glucose cotransporter-2.

[e]FAERS: Food and Drug Administration Adverse Event Reporting System.

[f]Control cohort.

[g]The total number of unique individual reports was used for the analysis to avoid overcounting cases that had more than 1 AE of interest listed.

Statistical Analysis

Descriptive Statistics

Frequencies for each AE PT code were calculated by the following equation:

$$\text{Frequency}(\%)=(\text{Number of reports with AE in a cohort})\times 100/\text{Number of reports in a cohort}$$

Comparative Statistics

AE report rates were compared via the ROR analysis using the following equations:

$$\text{ROR}=(a \div b)(c \div d)$$

$$\text{LnROR}=\text{Ln}(\text{ROR})$$

$$\text{SELnROR}=1/a+1/b+1/c+1/d$$

$$95\%\text{CI}=[\exp(\text{LnROR}-1.96\times\text{SELnROR}),\exp(\text{LnROR}+1.96\times\text{SELnROR})]$$

where a is the number of AE cases in the exposed group, b is the number of non-AE cases in the exposed group, c is the number of AE cases in the control group, and d is the number of non-AE cases in the control group.

Results

A total of 31,755 monotherapy reports, including 17,653 GLP-1 RA and 14,102 SGLT-2 inhibitor reports, were used for the analysis of 191 unique thyroid hyperplasia or neoplasm reports in the GLP-1 RA group and 7 reports in the SGLT-2 group, respectively (Tables 1 and 2).

GLP-1 RA monotherapy reports showed a statistically significant increase in thyroid hyperplasia and neoplasm AEs, with the ROR being 22.02 (95% CI 10.36-46.84) when compared to SGLT-2 inhibitors. When analyzed individually, every GLP-1 RA had significantly increased reporting of thyroid hyperplasia or neoplasm–related AEs when compared to SGLT-2 inhibitors. The results for all the GLP-1 RAs showed that the lower bound of the 95% CI of the ROR range was above

3: semaglutide (ROR 20.78, 95% CI 9.19-47.03), dulaglutide (ROR 17.79, 95% CI 7.86-40.25), exenatide (ROR 7.65, 95% CI 3.17-18.45), and liraglutide (ROR 35.92, 95% CI 16.71-77.20; Figure 1).

Table . Number of AE[a] reports and the respective demographics.

Demographics		Drug cohort				
		Semaglutide (n=3230)	Dulaglutide (n=3768)	Exenatide (n=4493)	Liraglutide (n=6162)	SGLT-2[b] inhibitors (canagliflozin, empagliflozin, and dapagliflozin; control[c]; n=14,102)
Unique individual thyroid hyperplasia or neoplasm reports, n (%)		33 (1.02)	33 (0.88)	17 (0.38)	108 (1.75)	7 (0.05)
Sex, n (%)						
	Male	4 (0.12)	14 (0.37)	3 (0.07)	35 (0.57)	3 (0.02)
	Female	27 (0.84)	14 (0.37)	13 (0.29)	66 (1.07)	2 (0.01)
	Unknown	2 (0.06)	5 (0.13)	1 (0.02)	7 (0.11)	2 (0.01)
Age (y)						
	Median	56	60	61	55	N/A[d]
	Mean (SD)	55.94 (13.41)	57.33 (11.35)	64.67 (6.35)	53.07 (11.08)	N/A

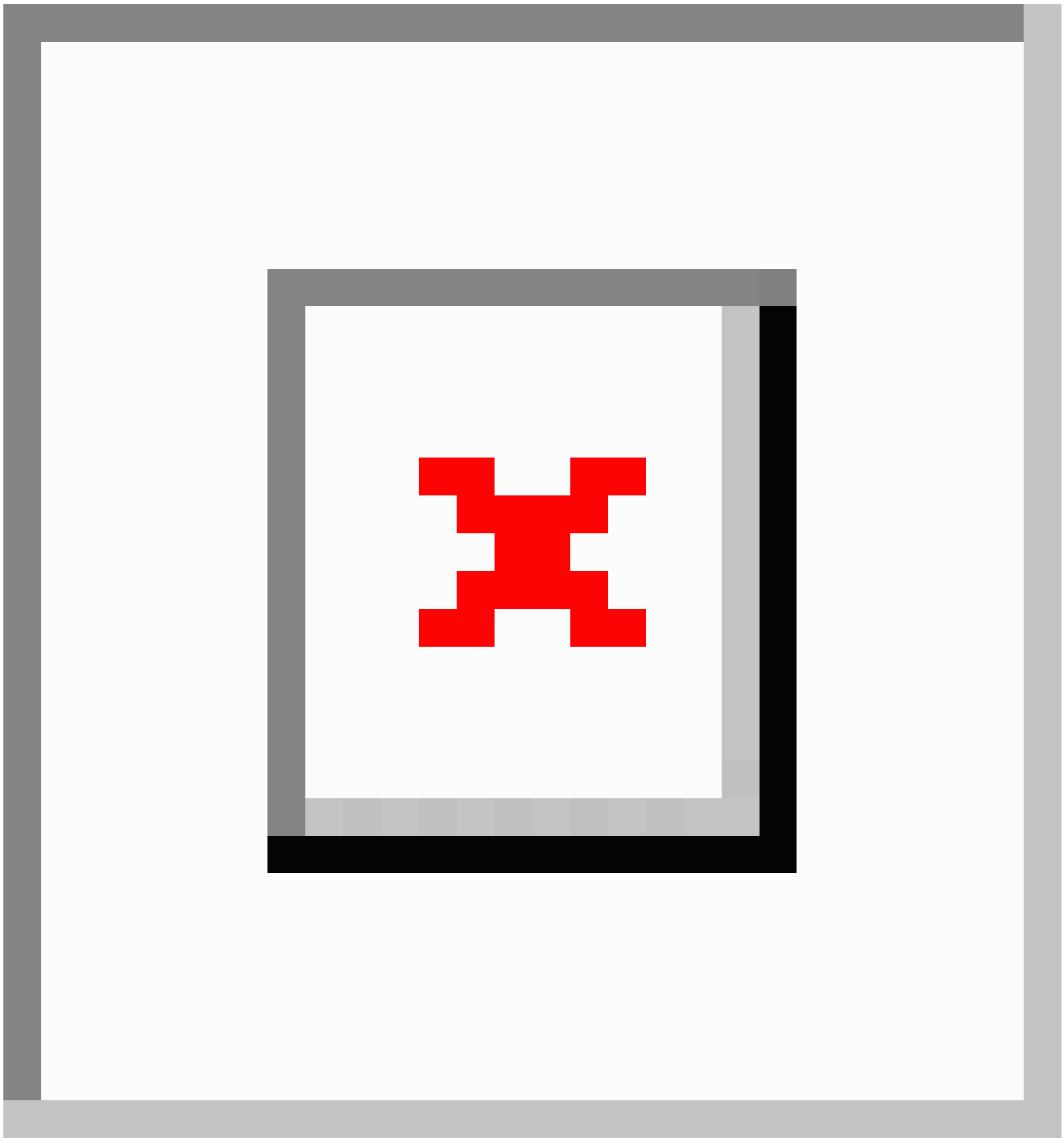
[a]AE: adverse event.

[b]SGLT-2: sodium-glucose cotransporter-2.

[c]Control cohort.

[d]N/A: not applicable.

Figure 1. Reporting odds ratios (RORs) of thyroid hyperplasia or neoplasm–related AEs in individual GLP-1 RA monotherapy cohorts and GLP-1 RAs as a class, when compared to the SGLT-2 inhibitor monotherapy control cohort. The x-axis is presented in logarithmic scale. AE: adverse event; GLP-1: glucagon-like peptide-1; nAE: number of adverse events; RA: receptor agonist; SGLT-2: sodium-glucose cotransporter-2.



Discussion

Principal Findings

In this study, we observed a reported association of GLP-1 RA treatment with thyroid hyperplasia and neoplasm AEs. To our knowledge, this is the first analysis of the FAERS to generate an ROR profile of GLP-1 RA monotherapy as a class and further analyze the individual GLP-1 RA monotherapy AEs compared to SGLT-2 inhibitors monotherapy AEs. The mean ages of the specific drug cohorts with these AEs were similar for comparison, ranging from 53.0 to 64.7 years. The thyroid hyperplasia and neoplasm AE association was significant for every GLP-1 RA in the class, with an even narrower range of 95% CI in the combined cohort. The RORs ranged from 7.65 to 35.92, with the lowest bound being 3.17. The highest mean ROR value was observed for liraglutide, and the lowest mean ROR was observed for exenatide, with only a small overlap in the 95% CIs. Interestingly, this trend was also seen in a similar analysis performed by Mali et al [20] using the European pharmacovigilance database (EudraVigilance), where this association was the strongest in the liraglutide cohort, followed by the exenatide cohort, with proportional reporting ranges of 27.5 (95% CI 22.7-33.3) and 22.5 (95% CI 17.9-28.3), respectively. The differences in the numbers and 95% CI ranges are due to the numbers of reports and type of analysis.

An association between T2DM and thyroid function has been previously established [28]. These disease states are often comorbid, and one affects the disease progression of the other [29]. However, the potential molecular mechanisms responsible for thyroid effects of GLP-1 RAs are insufficiently characterized and may include multiple pathways such as phosphoinositol-3 kinase and Akt serine/threonine kinase pathways; mitogen-activated protein kinase and extracellular signal-regulated kinase pathways; expression of GLP-1 receptors by C-cells; and GLP-1 association with triiodothyronine levels [21,30]. Thyroid hormone plays a pivotal role in nearly every aspect of lipid metabolism [31,32]. Thus, it was expected to observe thyroid hyperplasia and neoplasm–related AEs across all the T2DM drug cohorts that were investigated. However, only a single report of thyroid hyperplasia and neoplasm AE (thyroid cancer) was observed in the metformin monotherapy cohort, and only 7 were observed in the SGLT-2 inhibitor monotherapy cohort that was selected as the control due to similar T2DM disease-stage treatment guideline recommendations. In contrast, almost 200 of these AEs were reported for GLP-1 RAs in similarly sized cohorts, resulting in a statistically significant reported association. Therefore, as the GLP-1 RAs expand into non-T2DM indications such as obesity metabolic syndrome and other related conditions, controlled studies and better understanding of the molecular mechanisms of action are necessary to investigate this association and prevent potential serious consequences.

Study Limitations

Since this is an association study, the causality between the thyroid hyperplasias and neoplasms and GLP-1 RAs was not clinically adjudicated. However, this analysis of a large population–scale AE database provides a strong signal that may be clinically significant. The numbers of AEs presented in the study do not represent all treated patients due to voluntary submissions resulting in over- and underreporting [33,34]. Additionally, due the nature and long progression of the studied AEs, often requiring invasive procedures for a proper diagnosis, the cases may be significantly underreported. The case narratives with additional information such as a thorough medical history and test or diagnostic results were not available in the data sets provided by the FDA due to privacy concerns. Other limitations include the possible over-the-counter medications and supplements that are often not reported to health care professionals and may potentially add noise to the cohort compositions, AE frequencies, and to a lesser extent, RORs and the respective CIs. However, considering the large number of individuals in the GLP-1 RA and control cohorts, which were matched by indication, the signal was statistically significant.

Acknowledgments

We thank Dr Vahe Melkonyan, MD, for helpful discussions. We also that Dr Da Shi and Chris Edwards for help with data processing.

Data Availability

The data sets generated or analyzed during this study are available in the Food and Drug Administration Adverse Event Reporting System (FAERS) repository [24].

Authors' Contributions

HJ performed the experiments. RA and TM designed the study. RA, HJ, and TM drafted the manuscript and reviewed the final version. RA processed the data set.

Conflicts of Interest

None declared.

References

1. Ahmann AJ, Capehorn M, Charpentier G, et al. Efficacy and safety of once-weekly semaglutide versus exenatide ER in subjects with type 2 diabetes (SUSTAIN 3): a 56-week, open-label, randomized clinical trial. Diabetes Care 2018 Feb;41(2):258-266. [doi: 10.2337/dc17-0417] [Medline: 29246950]
2. Sorli C, Harashima SI, Tsoukas GM, et al. Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): a double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. Lancet Diabetes Endocrinol 2017 Apr;5(4):251-260. [doi: 10.1016/S2213-8587(17)30013-X] [Medline: 28110911]
3. Wysham C, Blevins T, Arakaki R, et al. Efficacy and safety of dulaglutide added onto pioglitazone and metformin versus exenatide in type 2 diabetes in a randomized controlled trial (AWARD-1). Diabetes Care 2014 Aug 1;37(8):2159-2167. [doi: 10.2337/dc13-2760] [Medline: 24879836]
4. Garber A, Henry R, Ratner R, et al. Liraglutide versus glimepiride monotherapy for type 2 diabetes (LEAD-3 Mono): a randomised, 52-week, phase III, double-blind, parallel-treatment trial. Lancet 2009 Feb 7;373(9662):473-481. [doi: 10.1016/S0140-6736(08)61246-5] [Medline: 18819705]

5. Buse JB, Bergenstal RM, Glass LC, et al. Use of twice-daily exenatide in basal insulin-treated patients with type 2 diabetes: a randomized, controlled trial. Ann Intern Med 2011 Jan 18;154(2):103-112. [doi: 10.7326/0003-4819-154-2-201101180-00300] [Medline: 21138825]
6. Garber AJ, Abrahamson MJ, Barzilay JI, et al. Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm--2016 executive summary. Endocr Pract 2016 Jan;22(1):84-113. [doi: 10.4158/EP151126.CS] [Medline: 26731084]
7. Blonde L, Umpierrez GE, Reddy SS, et al. American Association of Clinical Endocrinology clinical practice guideline: developing a diabetes mellitus comprehensive care plan-2022 update. Endocr Pract 2022 Oct;28(10):923-1049. [doi: 10.1016/j.eprac.2022.08.002] [Medline: 35963508]
8. Muzurović EM, Volčanšek Š, Tomšić KZ, et al. Glucagon-like peptide-1 receptor agonists and dual glucose-dependent insulinotropic polypeptide/glucagon-like peptide-1 receptor agonists in the treatment of obesity/metabolic syndrome, prediabetes/diabetes and non-alcoholic fatty liver disease-current evidence. J Cardiovasc Pharmacol Ther 2022 Jan;27:10742484221146371. [doi: 10.1177/10742484221146371] [Medline: 36546652]
9. Vilsbøll T, Christensen M, Junker AE, Knop FK, Gluud LL. Effects of glucagon-like peptide-1 receptor agonists on weight loss: systematic review and meta-analyses of randomised controlled trials. BMJ 2012 Jan 10;344:d7771. [doi: 10.1136/bmj.d7771] [Medline: 22236411]
10. Shaefer CJ, Kushner P, Aguilar R. User's guide to mechanism of action and clinical use of GLP-1 receptor agonists. Postgrad Med 2015;127(8):818-826. [doi: 10.1080/00325481.2015.1090295] [Medline: 26371721]
11. Drucker DJ. GLP-1 physiology informs the pharmacotherapy of obesity. Mol Metab 2022 Mar;57:101351. [doi: 10.1016/j.molmet.2021.101351] [Medline: 34626851]
12. Zhang F, Tong Y, Su N, et al. Weight loss effect of glucagon-like peptide-1 mimetics on obese/overweight adults without diabetes: a systematic review and meta-analysis of randomized controlled trials. J Diabetes 2015 May;7(3):329-339. [doi: 10.1111/1753-0407.12198] [Medline: 25043423]
13. Rubino DM, Greenway FL, Khalid U, et al. Effect of weekly subcutaneous semaglutide vs daily liraglutide on body weight in adults with overweight or obesity without diabetes: the STEP 8 randomized clinical trial. JAMA 2022 Jan 11;327(2):138-150. [doi: 10.1001/jama.2021.23619] [Medline: 35015037]
14. Dushay J, Gao C, Gopalakrishnan GS, et al. Short-term exenatide treatment leads to significant weight loss in a subset of obese women without diabetes. Diabetes Care 2012 Jan;35(1):4-11. [doi: 10.2337/dc11-0931] [Medline: 22040840]
15. Bjerre Knudsen L, Madsen LW, Andersen S, et al. Glucagon-like peptide-1 receptor agonists activate rodent thyroid C-cells causing calcitonin release and C-cell proliferation. Endocrinology 2010 Apr;151(4):1473-1486. [doi: 10.1210/en.2009-1272] [Medline: 20203154]
16. Madsen LW, Knauf JA, Gotfredsen C, et al. GLP-1 receptor agonists and the thyroid: C-cell effects in mice are mediated via the GLP-1 receptor and not associated with RET activation. Endocrinology 2012 Mar;153(3):1538-1547. [doi: 10.1210/en.2011-1864] [Medline: 22234463]
17. Byrd RA, Sorden SD, Ryan T, et al. Chronic toxicity and carcinogenicity studies of the long-acting GLP-1 receptor agonist dulaglutide in rodents. Endocrinology 2015 Jul;156(7):2417-2428. [doi: 10.1210/en.2014-1722] [Medline: 25860029]
18. Bezin J, Gouverneur A, Pénichon M, et al. GLP-1 receptor agonists and the risk of thyroid cancer. Diabetes Care 2023 Feb 1;46(2):384-390. [doi: 10.2337/dc22-1148] [Medline: 36356111]
19. Yang Z, Lv Y, Yu M, et al. GLP-1 receptor agonist-associated tumor adverse events: a real-world study from 2004 to 2021 based on FAERS. Front Pharmacol 2022 Oct 25;13:925377. [doi: 10.3389/fphar.2022.925377] [Medline: 36386208]
20. Mali G, Ahuja V, Dubey K. Glucagon-like peptide-1 analogues and thyroid cancer: an analysis of cases reported in the European pharmacovigilance database. J Clin Pharm Ther 2021 Feb;46(1):99-105. [doi: 10.1111/jcpt.13259] [Medline: 32926446]
21. Hu W, Song R, Cheng R, et al. Use of GLP-1 receptor agonists and occurrence of thyroid disorders: a meta-analysis of randomized controlled trials. Front Endocrinol (Lausanne) 2022 Jul 11;13:927859. [doi: 10.3389/fendo.2022.927859] [Medline: 35898463]
22. Craigle V. MedWatch: the FDA safety information and adverse event reporting program. J Med Libr Assoc 2007 Apr;95(2):224-225 [FREE Full text]
23. Kessler DA. Introducing MEDWatch. A new approach to reporting medication and device adverse effects and product problems. JAMA 1993 Jun 2;269(21):2765-2768. [doi: 10.1001/jama.269.21.2765] [Medline: 8492403]
24. FAERS data sets. Food and Drug Administration. URL: https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html [accessed 2024-04-15]
25. Keshishi D, Makunts T, Abagyan R. Common osteoporosis drug associated with increased rates of depression and anxiety. Sci Rep 2021 Dec 14;11(1):23956. [doi: 10.1038/s41598-021-03214-x] [Medline: 34907232]
26. Wollmer MA, Makunts T, Krüger THC, Abagyan R. Postmarketing safety surveillance data reveals protective effects of botulinum toxin injections against incident anxiety. Sci Rep 2021 Dec 21;11(1):24173. [doi: 10.1038/s41598-021-03713-x] [Medline: 34934096]
27. Makunts T, Jerome L, Abagyan R, de Boer A. Reported cases of serotonin syndrome in MDMA users in FAERS database. Front Psychiatry 2021 Jan 24;12:824288. [doi: 10.3389/fpsyt.2021.824288] [Medline: 35140642]

28. Kadiyala R, Peter R, Okosieme OE. Thyroid dysfunction in patients with diabetes: clinical implications and screening strategies. Int J Clin Pract 2010 Jul;64(8):1130-1139. [doi: 10.1111/j.1742-1241.2010.02376.x] [Medline: 20642711]
29. Jali MV, Kambar S, Jali SM, Pawar N, Nalawade P. Prevalence of thyroid dysfunction among type 2 diabetes mellitus patients. Diabetes Metab Syndr 2017 Nov;11 Suppl 1:S105-S108. [doi: 10.1016/j.dsx.2016.12.017] [Medline: 28057505]
30. Waser B, Blank A, Karamitopoulou E, Perren A, Reubi JC. Glucagon-like-peptide-1 receptor expression in normal and diseased human thyroid and pancreas. Mod Pathol 2015 Mar;28(3):391-402. [doi: 10.1038/modpathol.2014.113] [Medline: 25216224]
31. Duntas LH, Brenta G. A renewed focus on the association between thyroid hormones and lipid metabolism. Front Endocrinol (Lausanne) 2018 Sep 3;9:511. [doi: 10.3389/fendo.2018.00511] [Medline: 30233497]
32. Pucci E, Chiovato L, Pinchera A. Thyroid and lipid metabolism. Int J Obes Relat Metab Disord 2000 Jun;24 Suppl 2:S109-S112. [doi: 10.1038/sj.ijo.0801292] [Medline: 10997623]
33. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). Expert Opin Drug Saf 2017 Jul;16(7):761-767. [doi: 10.1080/14740338.2017.1323867] [Medline: 28447485]
34. Maciejewski M, Lounkine E, Whitebread S, et al. Reverse translation of adverse event reports paves the way for de-risking preclinical off-targets. Elife 2017 Aug 8;6:e25818. [doi: 10.7554/eLife.25818] [Medline: 28786378]

Abbreviations

AE: adverse event
FAERS: Food and Drug Administration Adverse Event Reporting System
FDA: Food and Drug Administration
GLP-1: glucagon-like peptide-1
PT: Preferred Term
RA: receptor agonist
ROR: reporting odds ratio
SGLT-2: sodium-glucose cotransporter-2
T2DM: type 2 diabetes mellitus

Edited by E Meinert, F Wu, T Leung; submitted 01.01.24; peer-reviewed by Anonymous; revised version received 03.03.24; accepted 10.03.24; published 01.05.24.

Please cite as:
Makunts T, Joulfayan H, Abagyan R
Thyroid Hyperplasia and Neoplasm Adverse Events Associated With Glucagon-Like Peptide-1 Receptor Agonists in the Food and Drug Administration Adverse Event Reporting System: Retrospective Analysis
JMIRx Med 2024;5:e55976
URL: https://xmed.jmir.org/2024/1/e55976
doi:10.2196/55976

Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study

Mara Giavina-Bianchi, MD, PhD; Edson Amaro Jr, MD, PhD; Birajara Soares Machado, PhD

Big Data Department, Hospital Israelita Albert Einstein, Sao Paulo, Brazil

Corresponding Author:
Mara Giavina-Bianchi, MD, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/50803

Companion article: https://med.jmirx.org/2024/1/e56496

Companion article: https://med.jmirx.org/2024/1/e56446

Companion article: https://med.jmirx.org/2024/1/e56441

Abstract

Background: The use of artificial intelligence (AI) in medicine has been a trending subject in the past few years. Although not frequently used in daily practice yet, it brings along many expectations, doubts, and fears for physicians. Surveys can be used to help understand this situation.

Objective: This study aimed to explore the degree of knowledge, expectations, and fears on possible AI use by physicians in daily practice, according to sex and time since graduation.

Methods: An electronic survey was sent to physicians of a large hospital in Brazil, from August to September 2022.

Results: A total of 164 physicians responded to our survey. Overall, 54.3% (89/164) of physicians considered themselves to have an intermediate knowledge of AI, and 78.5% (128/163) believed that AI should be regulated by a governmental agency. If AI solutions were reliable, fast, and available, 77.9% (127/163) intended to frequently or always use AI for diagnosis (143/164, 87.2%), management (140/164, 85.4%), or exams interpretation (150/164, 91.5%), but their approvals for AI when used by other health professionals (85/163, 52.1%) or directly by patients (82/162, 50.6%) were not as high. The main benefit would be increasing the speed for diagnosis and management (106/163, 61.3%), and the worst issue would be to over rely on AI and lose medical skills (118/163, 72.4%). Physicians believed that AI would be useful (106/163, 65%), facilitate their work (140/153, 91.5%), not alter the number of appointments (80/162, 49.4%), not interfere in their financial gain (94/162, 58%), and not replace their jobs but be an additional source of information (104/162, 64.2%). In case of disagreement between AI and physicians, most (108/159, 67.9%) answered that a third opinion should be requested. Physicians with ≤10 years since graduation would adopt AI solutions more frequently than those with >20 years since graduation (P=.04), and female physicians were more receptive to other hospital staff using AI than male physicians (P=.008).

Conclusions: Physicians were shown to have good expectations regarding the use of AI in medicine when they apply it themselves, but not when used by others. They also intend to use it, as long as it was approved by a regulatory agency. Although there was hope for a beneficial impact of AI on health care, it also brings specific concerns.

(JMIRx Med 2024;5:e50803) doi:10.2196/50803

KEYWORDS

artificial intelligence; adoption; AI; acceptance; opinion; perceptions; survey; expectations; physician; medical survey; qualitative study

Introduction

The use of artificial intelligence (AI) is expanding throughout the field of medicine, driven by researchers and entrepreneurs [1,2]. Over the last decade, the number of publications on AI in medicine and biomedicine has substantially increased [3]. AI solutions might change the clinical practice in nearly all medical disciplines and areas of health care. Despite the potential of machine learning to improve multiple aspects of

patient care, there are still barriers to clinical adoption. Important questions remain regarding how machine learning interventions are being incorporated into health care [4]. A reluctance to adopt AI-based solutions might be due to a lack of knowledge; fear of error; and concerns about losing jobs, power, or both [5]. Another perceived limitation of AI applications is the belief that communication and empathy are human competencies that cannot be replaced by AI. In addition, the ability to provide value-based care requires the physicians' judgments. Some possible benefits included expectations about improved efficiencies, especially with respect to the reduction of administrative burdens on physicians [6]. Examples of practical use of AI solutions in clinical routine are still scarce around the globe [2]. The increasing development in medical systems using AI brings enormous expectations and fears for both physicians and patients.

Physicians are likely to be the "earliest" adopters of AI solutions for patient care and inevitably should become direct AI operators. Therefore, they play a pivotal role in the acceptance and implementation of clinical AI, and consequently, their views need to be known, explored, and understood [7]. Opinion surveys are important tools in assessing satisfaction with a particular service and consist of a list of questions whose objective is to extract certain data from a group of people [8]. Previous studies on the acceptance of the use of AI in medicine were limited to specific areas, such as radiology [5,9-11], dermatology [1,12-14], and ophthalmology [15-17], as well as to specific countries. However, at the time of writing, we were unable to find studies exploring this subject among Brazilian physicians, and there are very few studies in Latin America, leaving a gap in this part of the globe. Is the perception of AI adoption similar to other countries? At the same time, some aspects are not yet explored—does the acceptance of AI solutions vary according to the number of years since medical graduation or the physician's sex? Routine observation could indicate that younger individuals are keener to accept new technologies, and no information regarding sex preferences on this matter has yet been evaluated, as far as we know.

The main objective of this study was to assess the expectations, fears, and thoughts of physicians from a single-center private hospital about some practical aspects of the hypothetical use of AI solutions in medical daily practice. The secondary objective was to verify if there were differences in the opinions among physicians according to the sex and time since graduation.

Methods

Ethical Considerations

This study was approved by the Ethics Committee of Hospital Israelita Albert Einstein (CAAE: 30749620.6.0000.0071). All participants provided informed consent, and the collected data were anonymized. There were no financial incentives for answering the questionnaire.

Study Design and Target Population

We performed a cross-section observational study via an opinion survey. Our target population was physicians who were part of the clinical staff of the Hospital Israelita Albert Einstein (HIAE). The survey was sent to 7457 physicians from the clinical staff with an email address registered in the HIAE marketing department. There were no exclusion criteria, and responses were collected from August 1 to October 1, 2022 (60 days). The HIAE is a nonprofit, public interest organization based in Sa˜o Paulo, Brazil. Although physicians of the HIAE do not represent the entire population of physicians in Brazil, their answers can give some insights on the subject, since, at this moment, we have none.

Questionnaire

The questionnaire was composed of 30 questions. Question 1 was the informed consent form (ICF). In case of acceptance of the ICF, the next questions were presented to the individual. If not, the survey would be terminated. The time to complete the questionnaire was approximately 7 minutes on average. The survey was completely anonymous and confidential, and only the authors of this work had access to the answers. The complete questionnaire can be viewed in Multimedia Appendix 1. The questions were developed by the authors based on their experience in developing AI solutions for physicians and on previous published medical literature.

No specific technology was evaluated. Most of the questions asked about the use of AI algorithms for the diagnosis or management of diseases, aiming to capture the possible expectations of our target population of physicians in clinical practice. The questionnaire was divided into 5 sections. Section 1 was the ICF (question 1). Section 2 (questions 2-12) was designed to profile the physicians, including sex; age; highest level of education; medical specialty; years since graduation; private versus public sector work; city and state of work; self-assessment knowledge of AI in general (not specific for health care AI solutions); and the use of computer or smartphone applications that use AI solutions for daily tasks, such as WhatsApp, Instagram, Facebook, Waze, Google Maps, and bank apps, among others. We did not ask specifically about the use of AI solutions for health care in their daily work, only if they were aware of AI solutions in medicine. Section 3 (questions 13-18) explored the physicians' thoughts about AI solutions for diagnosis, management, subsidiary exams, and interpretation of diseases, such as COVID-19, and about the use of AI solutions for the diagnosis or treatment of diseases by nurses or physiotherapists or directly by the patient. We also proposed a hypothetical exercise to evaluate physicians' level of anxiety and actions taken if they had received a suspicious diagnosis of melanoma for 1 of their skin lesions by an AI algorithm. Section 4 (questions 19-24) asked about expected benefits and problems, possible frequency of AI adoption, workload, and utility. Section 5 (questions 25-30) were about physicians' replacement by AI solutions, financial expectations, possible scenarios of disagreements between AI and physicians, and legal and regulatory aspects. Along with the questions, there were many opportunities for physicians to provide comments about the answers in an open-text box. Physicians could skip any question; thus, the number of responders could vary among the questions.

Emailing the Questionnaire

The survey was sent by email to all physicians with an email address linked to the HIAE. In the first email, sent on August 1, 2022, a brief introduction inviting the physician to participate in the survey and the questionnaire link to be completed in the SurveyMonkey computer program (SurveyMonkey Inc) [18] were sent to all physicians. In the second email, sent on August 8, 2022, we replicated the same email to those who had not completed the survey from the previous email. There was a third email sent on September 5, 2022, to the remaining individuals. The survey ended on September 30, 2022. This time frame of 60 days and the number of reminders during the period followed the current modus operandi of the hospital's marketing department for all studies sent electronically. SurveyMonkey's program has a blocking mechanism that prevents the same participant from responding to the survey more than once. It identifies and notifies the user that the questionnaire had already been answered, blocking a new response. The survey was previously tested by 3 physicians of the HIAE medical team who were part of the target population. Our work followed the CROSS (Checklist for Reporting Survey Studies) [4].

Statistical Analysis

The response rate was the number of physicians who responded the questionnaire divided by the number of physicians to whom the email was sent multiplied by 100. The completion rate was the number of surveys answered and sent divided by the number of surveys initiated by respondents multiplied by 100. There were 2 different questions involving statistical analysis. In the first question—"does the time since medical graduation matter?"—the participants were divided into 3 groups (≤10 years, 11-20 years, or >20 years) according to their answer to question 7 of the questionnaire. In the second question—"does the subject's sex matter?"—physicians were divided into male or female individuals according to their answer to question 3. Statistical analyses between for both analyses were performed using the χ^2 test in Prism software (version 6; GraphPad Software Inc). *P* values <.05 were considered significant.

Results

Participant Characteristics

In all, 181 physicians accepted the ICF. The response rate was 2.4% (181/7457). The completion rate of the questionnaire was 90.6% (164/181). As physicians could skip any question, the number of responders could vary among the questions.

Table 1 shows the profile of the physicians who responded to the survey. They were mostly male (111/171, 64.9%), 36-55 years of age (99/171, 57.9%), and with >20 years since medical graduation (110/171, 64.3%). As for the place of work, 68.4% (117/171) worked mainly in the private sector, 26.9% (46/171) worked equally in both sectors, and 5.3% (9/171) worked mainly in the public sector. All academic degrees were present in the study; most had a residency or specialization internship (57/171, 33.3%), doctorate degree (55/171, 32.2%), or master's degree (43/171, 25.1%). The distribution among specialties was heterogeneous and skewed: surgeons were the most frequent (37/169, 21.9%). This finding may reflect the different number of physicians of each specialty linked to the hospital mailing list. There were probably many more pediatricians in the hospital's mailing list than psychiatrists. Almost all (166/171, 97%) of them worked in São Paulo city, as the main hospital is in this location. Most (89/164, 54.3%) of them classified their own knowledge of AI as intermediate. Most of the participants used smartphones or computers applications that incorporate AI algorithms for daily tasks outside of work (119/164, 72.6%) and claimed to be aware of AI algorithms applied specifically to medicine (86/164, 52.4%). There were no statistical differences among the participants based on sex. For the group**s based on** "years since **graduation,**" there was a significant *P* value (*P*=.049) in the self-assessment of AI knowledge.

Table . Profile of physicians who answered the opinion questionnaire on artificial intelligence (AI) solutions at the Hospital Israelita Albert Einstein (questions 2-12).

Questions and responses		Years since graduation, n (%)			Sex, n (%)		Total (n=171), n (%)
		≤10 (n=13)	11-20 (n=48)	>20 (n=110)	Male (n=111)	Female (n=60)	
2. Sex							
	Female	6 (46)	18 (38)	36 (33)	0 (0)	60 (100)	60 (35)
	Male	7 (54)	30 (63)	74 (67)	111 (100)	0 (0)	111 (65)
3. Age range (years)							
	26-35	11 (85)	2 (4)	0 (0)	9 (8)	4 (7)	13 (8)
	36-45	2 (15)	43 (90)	3 (3)	29 (26)	19 (32)	48 (28)
	46-55	0 (0)	3 (6)	48 (44)	29 (26)	22 (37)	51 (30)
	56-65	0 (0)	0 (0)	33 (30)	21 (19)	12 (20)	33 (19)
	>65	0 (0)	0 (0)	26 (24)	23 (21)	3 (5)	26 (15)
4. Highest academic degree (>20 years since graduation: n=111)							
	Medical	0 (0)	0 (0)	1 (1)	1 (1)	0 (0)	1 (1)
	Residency or specialization internship	10 (77)	25 (52)	22 (20)	35 (32)	22 (37)	57 (33)
	Master's degree	3 (23)	13 (27)	27 (24)	27 (24)	16 (27)	43 (25)
	Doctorate degree	0 (0)	7 (15)	48 (43)	36 (32)	18 (30)	54 (32)
	Postdoctoral research	0 (0)	3 (6)	6 (5)	6 (5)	3 (5)	9 (5)
	Associated professor	0 (0)	0 (0)	6 (5)	6 (5)	0 (0)	6 (4)
	Other	0 (0)	0 (0)	1 (1)	0 (0)	1 (2)	1 (1)
5. Specialty (>20 years since graduation: n=108; male: n=57; female: n=168)							
	Dermatology	0 (0)	0 (0)	3 (3)	0 (0)	3 (5)	3 (2)
	Gynecology or obstetrics	0 (0)	1 (2)	16 (15)	9 (8)	8 (14)	17 (10)
	Internal medicine	2 (2)	11 (23)	12 (11)	14 (13)	11 (19)	25 (15)
	Management	1 (17)	4 (8)	1 (1)	5 (5)	1 (2)	6 (4)
	Ophthalmology	0 (0)	0 (0)	3 (3)	3 (3)	0 (0)	3 (2)
	Orthopedics	0 (0)	3 (6)	7 (6)	10 (9)	0 (0)	10 (6)
	Otorhinolaryngology	0 (0)	0 (0)	5 (5)	4 (4)	1 (2)	5 (3)
	Other	4 (33)	12 (25)	17 (16)	18 (16)	14 (24)	32 (19)
	Pathology	0 (0)	0 (0)	2 (2)	1 (1)	1 (1)	2 (1)
	Pediatrics	4 (33)	6 (0)	12 (11)	10 (9)	12 (20)	22 (13)
	Psychiatry	0 (0)	0 (0)	2 (2)	0 (0)	2 (3)	2 (1)
	Radiology	0 (0)	3 (6)	1 (1)	2 (2)	2 (3)	4 (2)
	Surgery	2 (17)	8 (17)	27 (25)	35 (32)	2 (7)	37 (22)
6. Years since graduation							
	5-10	13 (100)	0 (0)	0 (0)	7 (6)	6 (10)	13 (8)
	11-20	0 (0)	48 (100)	0 (0)	30 (27)	18 (30)	48 (28)
	>20	0 (0)	0 (0)	110 (100)	74 (67)	36 (60)	110 (64)

Questions and responses		Years since graduation, n (%)			Sex, n (%)		Total (n=171), n (%)
		≤10 (n=13)	11-20 (n=48)	>20 (n=110)	Male (n=111)	Female (n=60)	
7. Work sector							
	Mostly public	1 (1)	3 (63)	5 (5)	4 (4)	5 (8)	9 (5)
	Mostly private	8 (62)	35 (73)	73 (66)	77 (69)	39 (65)	116 (68)
	Equally in both	4 (31)	10 (21)	32 (29)	30 (27)	16 (27)	46 (27)
8. Workplace (state; male: n=110; female: n=59; total: n=169)							
	São Paulo	13 (100)	48 (100)	108 (98)	108 (98)	59 (100)	167 (99)
	Other	0 (0)	0 (0)	2 (2)	2 (2)	0 (0)	2 (1)
9. Workplace (city; >20 years since graduation: n=111)							
	Capital	13 (100)	48 (100)	105 (95)	107 (96)	58 (97)	165 (96)
	Coast or inland	0 (0)	0 (0)	6 (5)	4 (4)	2 (3)	6 (4)
10. Self-assessment of the degree of AI knowledge in general (years since graduation, <10: n=12; 11-20: n=46; >20: n=106; male: n=108; female: n=55; total: n=163)[a]							
	Low	0 (0)	19 (41)	30 (28)	27 (25)	21 (38)	48 (29)
	Medium	8 (67)	20 (43)	61 (58)	60 (56)	29 (53)	89 (55)
	High	4 (33)	7 (15)	12 (11)	19 (18)	4 (7)	23 (14)
	None	0 (0)	0 (0)	3 (3)	2 (2)	1 (2)	3 (2)
11. Frequency of AI use for daily life activities (years since graduation, <10: n=12; 11-20: n=45; >20: n=103; male: n=105; female: n=53; total: n=158)							
	Never	0 (0)	0 (0)	1 (1)	0 (0)	1 (2)	2 (1)
	Rarely	1 (8)	1 (2)	2 (2)	3 (28)	1 (2)	4 (2)
	Sometimes	1 (8)	11 (24)	22 (21)	22 (20)	12 (22)	34 (21)
	Frequently	5 (42)	21 (46)	49 (46)	48 (44)	26 (47)	74 (45)
	Always	5 (42)	12 (26)	28 (26)	31 (29)	14 (25)	45 (27)
	Do not know	0 (0)	1 (2)	4 (4)	4 (4)	1 (2)	5 (3)
12. Aware of AI solutions in medicine? (years since graduation, <10: n=12; 11-20: n=46; >20: n=106; male: n=108; female: n=55; total: n=163)							
	Yes	11 (92)	24 (52)	51 (48)	56 (52)	29 (53)	85 (52)
	No	0 (0)	13 (28)	35 (33)	34 (31)	14 (25)	48 (29)
	Uncertain	1 (8)	9 (20)	20 (19)	18 (17)	12 (22)	30 (18)

[a]*P*=.049 when comparing groups based on years since graduation.

The Use of AI Solutions

In total, 164 participants answered all questions until the end of the questionnaire. Figure 1 summarizes the answers from questions 13-17 of the survey, and Table 2 shows the answers to questions 13-17 according to the studied groups. For physicians in general, there was a belief that AI algorithms would be helpful for patients' diagnosis (143/164, 87.2%) and management (140/164, 85.5%) and would support image exams interpretation (150/164, 91.5%) when used by them. They were divided about the use of AI by other health professionals, such as nurses or physiotherapists (85/163, 52.1% approved), or by the patient themselves (82/162, 50.6% approved).

Figure 1. Physicians' expectations about the role of artificial intelligence (AI) solutions when used by themselves, other health care professionals, or patients (questions 13-17).

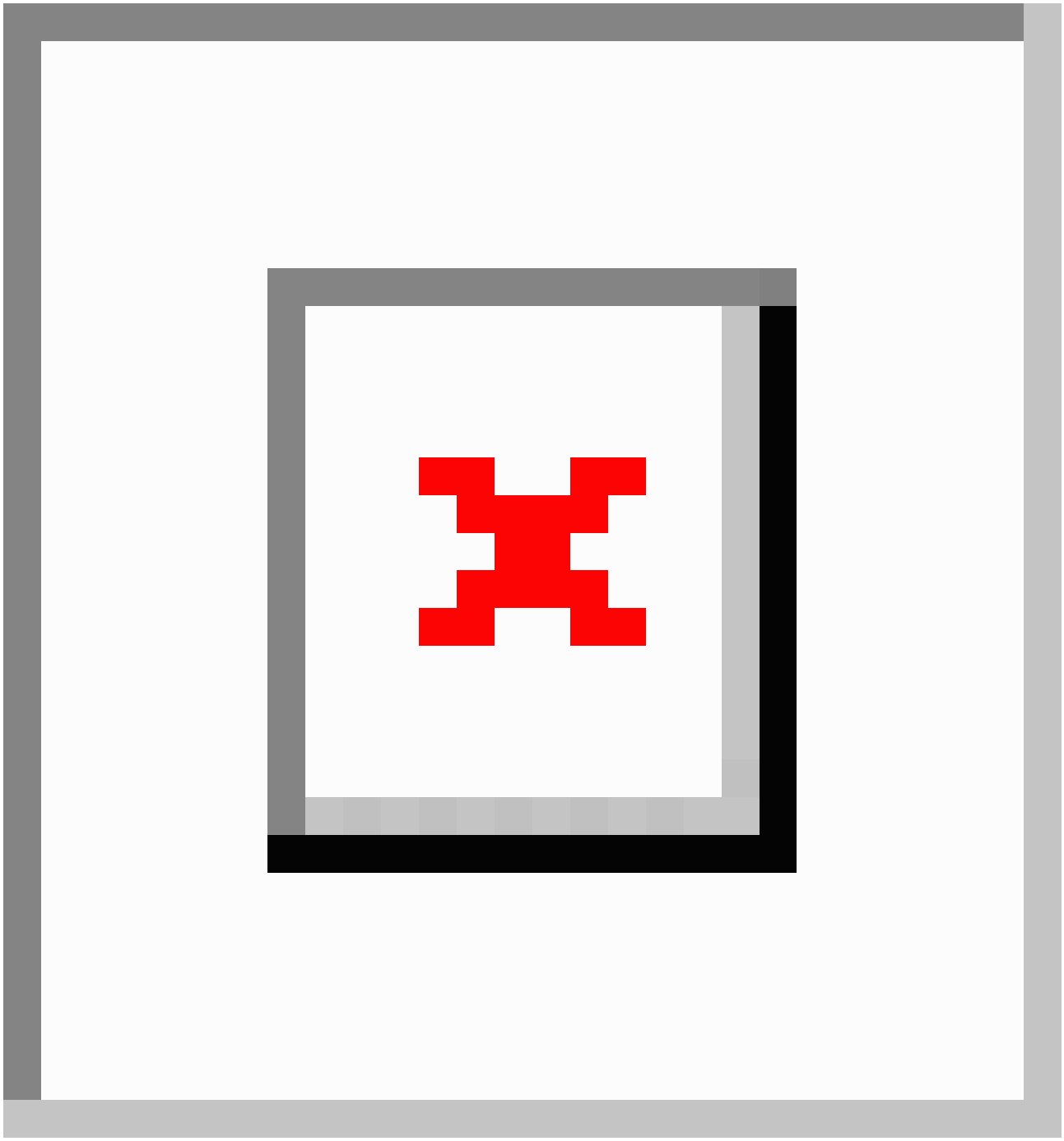


Table . Physicians' expectations about the role of artificial intelligence (AI) solutions when used by themselves, other health care professionals, or patients (questions 13-17), according to years since graduation and sex.

Questions and responses	Years since graduation n (%)			Sex n (%)	
	≤10 (n=12)	11-20 (n=46)	>20 (n=106)	Male (n=108)	Female (n=55)
13. Use of AI to support diagnosis by physicians (eg, COVID-19)					
Totally in favor	7 (58)	15 (33)	41 (39)	45 (42)	18 (33)
In favor	4 (33)	23 (50)	53 (50)	50 (46)	29 (53)
Not in favor nor against	0 (0)	8 (17)	11 (10)	11 (10)	8 (15)
Against	1 (8)	0 (0)	1 (1)	2 (2)	0 (0)
Totally against	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
14. Use of AI to support disease management by physicians (eg, corticosteroids for COVID-19)					
Totally in favor	5 (42)	13 (28)	33 (31)	37 (34)	14 (25)
In favor	6 (50)	22 (48)	57 (54)	54 (50)	30 (55)
Not in favor nor against	1 (8)	10 (22)	15 (14)	15 (14)	11 (20)
Against	0 (0)	1 (2)	1 (1)	2 (2)	0 (0)
Totally Against	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
15. Use of AI to support exam interpretation (eg, x-rays for SARS)					
Totally in favor	8 (67)	19 (41)	42 (40)	49 (45)	20 (36)
In favor	3 (25)	23 (50)	55 (52)	47 (44)	33 (60)
Not in favor nor against	0 (0)	4 (9)	8 (8)	10 (9)	2 (4)
Against	1 (8)	0 (0)	1 (1)	2 (2)	0 (0)
Totally against	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
16. Use of AI to aid diagnosis or management by other hospital staff (nurses or physiotherapists; >20 years since graduation: n=105; male: n=107)[a]					
Totally in favor	3 (25)	8 (17)	14 (13)	18 (17)	7 (13)
In favor	6 (50)	14 (30)	37 (35)	29 (27)	28 (51)
Not in favor nor against	2 (17)	10 (22)	22 (21)	21 (20)	12 (22)
Against	0 (0)	8 (17)	22 (21)	23 (21)	7 (13)
Totally against	1 (8)	6 (13)	10 (10)	16 (15)	1 (2)
17. Use of AI directly by patients to aid diagnosis or management (years since graduation, <10: n=10; 11-20: n=47; >20: n=107; male: n=107; female: n=54)					
Totally in favor	5 (42)	7 (15)	21 (21)	24 (22)	9 (17)
In favor	2 (17)	15 (33)	32 (31)	30 (28)	19 (35)
Not in favor nor against	2 (17)	10 (22)	22 (21)	15 (14)	16 (30)
Against	0 (0)	8 (17)	22 (21)	23 (21)	7 (13)
Totally against	1 (8)	7 (15)	10 (10)	15 (14)	3 (6)

[a]*P*=.008 when comparing male and female physicians.

Interestingly, there was a significant difference in the use of AI solutions by other hospital professionals according to sex: female physicians were more favorable toward it than male physicians (*P*=.008).

When placing themselves in the role of a patient, physicians acknowledged that AI diagnosis solutions when used by nonspecialists might cause distress about some types of diagnosis, such as skin melanoma (question 18). This question asked that if a melanoma detection solution was used by the physicians on themselves regarding a certain lesion and the AI

showed a high probability of melanoma diagnosis, what that situation would elicit in their feelings (the degree of anxiety or none) and what action would be taken (how quick they would seek a specialist appointment or if they would not seek one). Overall, 91% (142/156) reported that they would be at least anxious and seek an appointment as soon as possible. The major benefits cited by the physicians were greater speed for diagnosis and management (104/432, 24.1% of responses and 104/164, 63.4% of responders), greater accuracy (84/432, 19.4%), health care cost reduction (84/432, 19.4%), and greater access (67/432, 15.5%). The main issues listed were the fear of relying excessively on the AI algorithms and causing physicians to lose their medical skills (115/447, 25.7% of responses and 115/164, 70.1% of responders), wrongful diagnostic or management reports (84/447, 18.8%), and increasing the distance in the physician-patient relationship (84/447, 18.8%). The intention to adopt AI solutions showed a significant difference among the groups based on years since graduation (P=.04). Individuals with ≤10 years since graduation would use AI most of the times or always (10/12, 83%) in greater number than the other 2 groups: 30% (14/46) and 32.4% (34/105) for the groups with 11-20 and >20 years since graduation, respectively (Table 3). When the 3 groups were tested separately, 2 by 2, there was a significant P value of .004 when comparing ≤10 and >20 years since graduation and a P value of .02 when comparing ≤10 and 11-20 years since graduation, proving that the difference was between the group with ≤10 years since graduation and the other 2 groups. When comparing the groups with 11-20 and >20 years since graduation, the P value was .95 for that same question.

Table . Overview and expectations, frequency of adoption, work facilitation, number of appointments, and utility about artificial intelligence (AI) solutions according to years since graduation and sex (questions 18-24).

Questions and responses	Years since graduation, n (%)			Sex, n (%)	
	≤10 (n=12)	11-20 (n=46)	>20 (n=105)	Male (n=108)	Female (n=54)
18. Consequences of AI diagnosis if you were a patient (eg, AI solution made a diagnosis of suspicious melanoma for 1 of your skin lesions; years since graduation, 11-20: n=44; >20: n=104; male: n=106; female: n=53)					
Extremely anxious and immediate appointment	10 (83)	26 (59)	64 (62)	63 (59)	36 (68)
Anxious and appointment whenever possible	2 (17)	13 (30)	28 (77)	32 (30)	11 (21)
Not shaken and appointment whenever possible	0 (0)	4 (9)	9 (9)	11 (10)	2 (38)
Not shaken and no appointment	0 (0)	1 (2)	0 (0)	0 (0)	1 (2)
I am a dermatologist	0 (0)	0 (0)	3 (3)	0 (0)	3 (6)
19. Expected benefits of AI (physicians could pick up to 3 options; years since graduation, <10: n=33; 11-20: n=131; >20: n=285; male: n=291; female: n=141)					
Greater speed	9 (27)	33 (25)	64 (22)	67 (23)	37 (26)
Greater accuracy	6 (18)	27 (21)	54 (19)	58 (20)	26 (18)
Cost reduction	5 (15)	24 (18)	58 (20)	57 (20)	27 (19)
Reduction in the number of subsidiary exams	5 (15)	16 (12)	34 (12)	35 (12)	18 (13)
Reduction in patient anxiety	0 (0)	1 (1)	8 (3)	7 (2)	1 (1)
Greater access to health care	7 (21)	20 (15)	42 (15)	43 (15)	24 (17)
Greater patient participation in health care	1 (3)	8 (6)	22 (8)	20 (7)	8 (6)
Other	0 (0)	2 (2)	3 (1)	4 (14)	0 (0)
20. Expected problems of AI (physicians could pick up to 3 options; years since graduation, <10: n=33; 11-20: n=134; >20: n=295; male: n=305; female: n=142)					
Confidentiality issues	4 (12)	7 (5)	20 (7)	17 (6)	12 (8)
Worsening of the physician-patient relationship	4 (12)	23 (17)	59 (20)	57 (19)	27 (19)
Wrongful use of patient's information by employers and insurance companies	7 (21)	14 (10)	47 (16)	47 (15)	19 (13)
Errors in diagnosis or management	5 (15)	26 (19)	55 (19)	55 (18)	29 (20)
Physicians relying too much on AI and losing medical skills	9 (27)	35 (26)	74 (25)	78 (26)	37 (26)
Increase in health care cost	1 (3)	8 (6)	10 (3)	13 (4)	4 (3)

Questions and responses		Years since graduation, n (%)			Sex, n (%)	
		≤10 (n=12)	11-20 (n=46)	>20 (n=105)	Male (n=108)	Female (n=54)
	Lack of AI transparency	2 (6)	15 (11)	28 (9)	31 (10)	12 (8)
	Other	1 (3)	6 (4)	2 (1)	7 (2)	2 (1)
21. Expected frequency of AI adoption if the algorithm was reliable and only needed up to 2 min to provide an answer [a]						
	Never	0 (0)	0 (0)	1 (1)	1 (1)	0 (0)
	Rarely	1 (8)	1 (2)	1 (1)	3 (3)	0 (0)
	Sometimes	1 (8)	10 (22)	20 (19)	18 (17)	13 (24)
	Frequently	0 (0)	21 (46)	48 (46)	44 (41)	24 (44)
	Most of the times	4 (33)	6 (13)	17 (16)	19 (18)	8 (15)
	Always	6 (50)	8 (17)	17 (16)	22 (41)	9 (17)
	Do not know	0 (0)	0 (0)	1 (1)	1 (1)	0 (0)
22. Work facilitation by AI (assuming the conditions in question 21; years since graduation, <10: n=11; 11-20: n=73; >20: n=104; male: n=106)						
	Makes work easier	9 (82)	42 (91)	89 (86)	92 (87)	47 (87)
	Makes work more difficult	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Does not change	2 (18)	2 (4)	9 (9)	9 (8)	4 (7)
	Do not know	0 (0)	29 (4)	6 (6)	5 (5)	3 (6)
23. Number of daily appointments (assuming the conditions in question 21)						
	Increases	6 (50)	14 (30)	34 (32)	33 (31)	20 (37)
	Decreases	0 (0)	2 (4)	7 (7)	5 (5)	4 (7)
	Stays the same	5 (42)	24 (52)	51 (49)	55 (51)	25 (46)
	Do not know	1 (8)	6 (13)	13 (12)	15 (14)	5 (9)
24. AI utility in daily work (assuming the conditions in question 21)						
	Useful for diagnosis	2 (17)	15 (33)	21 (20)	25 (23)	13 (24)
	Useful for management	0 (0)	1 (2)	9 (9)	8 (7)	2 (4)
	Useful for diagnosis and management	9 (75)	27 (59)	70 (67)	68 (63)	38 (70)
	Not help nor hinder	0 (0)	0 (0)	2 (2)	2 (2)	0 (0)
	Hinders	1 (8)	1 (2)	0 (0)	2 (2)	0 (0)
	Do not know	0 (0)	2 (4)	3 (3)	3 (3)	1 (2)

[a] P=.04 when comparing groups based on years since graduation.

Overall, we can see in Table 3 that physicians intended to apply AI in medicine frequently (69/163, 42.3%) and believed that it would facilitate their work (140/153, 91.5%). Further, 49.1% (80/163) answered that AI would not interfere with the number of appointments, whereas one-third (54/163, 33.1%) of them believed that it would increase the number of appointments. They also believed that it would be useful for patients' diagnosis and management (106/163, 65%).

Table 4 details the answers for questions 25-30. The answers revealed that physicians perceived that AI algorithms would not replace them but rather be 1 more source of information (105/163, 64.4%) and would not alter their financial gain (94/163, 57.7%). In the event of a diagnosis or conduct disagreement between physicians and the AI solution, we proposed 2 scenarios. In the first scenario, AI algorithms and physicians had the same accuracy rate for a defined task, and in the second scenario, AI had a better accuracy rate than physicians. We asked the physicians what should be done in each case. In the former case, physicians were divided between "asking for a third opinion" (86/162, 53.1%) or that "the medical opinion should be followed" (72/162, 44.4%). In the latter case, the majority (108/160, 67.5%) chose to request a third opinion. As for legal responsibility, most individuals (88/159, 55.3%)

answered that it should be shared between the AI algorithm's manufacturer and the physicians and hospitals. Further, 78.5% (128/163) responded that AI solutions should have the stamp of a regulatory governmental agency. No statistical differences in answers were found between groups based on years since graduation or sex, but we can see a trend for female physicians being more favorable toward governmental regulation (*P*=.055; not statistically significant).

Table . Effects of artificial intelligence (AI) solutions on the routine of medical work among those who answered the opinion questionnaire at the Hospital Israelita Albert Einstein (questions 25-30).

Questions and responses		Years since graduation, n (%)			Sex, n (%)	
		≤10 (n=12)	11-20 (n=46)	>20 (n=105)	Male (n=108)	Female (n=54)
25. Physician's replacement in medical specialties based on imaging? (radiology, dermatology, pathology, etc)						
	Totally	1 (8)	2 (43)	2 (2)	5 (5)	0 (0)
	Partially	2 (17)	15 (33)	34 (32)	37 (34)	14 (26)
	One more source of information	9 (75)	28 (61)	68 (65)	64 (59)	40 (74)
	Not alter	0 (0)	1 (2)	0 (0)	1 (1)	0 (0)
	Do not know	0 (0)	0 (0)	1 (1)	1 (1)	0 (0)
26. Financial gain						
	Increases	5 (42)	5 (11)	12 (11)	15 (14)	7 (13)
	Decreases	1 (8)	7 (15)	18 (17)	19 (18)	7 (13)
	Not altered	6 (50)	27 (59)	61 (58)	61 (56)	33 (61)
	Do not know	0 (0)	7 (15)	14 (13)	13 (12)	7 (13)
27. What to do if there is a disagreement between AI and physicians (supposing accuracies are equal)? (>20 years since graduation: n=104; female: n=53)						
	Favors physicians' opinion	5 (42)	23 (50)	44 (42)	52 (48)	20 (38)
	Favors AI's opinion	0 (0)	0 (0)	1 (0)	1 (1)	0 (0)
	Request a third opinion	7 (58)	23 (50)	56 (54)	52 (48)	33 (61)
	Do not know	0 (0)	0 (0)	3 (3)	3 (3)	0 (0)
28. What to do if there is a disagreement between AI and physicians (supposing AI accuracy is greater than physicians')? (>20 years since graduation: n=102; male: n=107; female: n=52)						
	Favors physicians' opinion	4 (33)	9 (20)	16 (16)	22 (21)	7 (13)
	Favors AI's opinion	0 (0)	8 (17)	10 (10)	12 (11)	6 (12)
	Request a third opinion	8 (67)	29 (63)	71 (70)	69 (64)	39 (75)
	Do not know	0 (0)	0 (0)	5 (5)	4 (4)	0 (0)
29. Legal liability (>20 years since graduation: n=101; male: n=106; female: n=52)						
	Physician only	3 (25)	10 (22)	32 (32)	35 (33)	9 (17)
	AI only	1 (8)	1 (2)	6 (6)	5 (5)	3 (6)
	Shared equally	7 (58)	32 (70)	49 (49)	56 (53)	32 (62)
	Do not know	1 (8)	3 (7)	14 (14)	10 (9)	8 (15)
30. Governmental regulation [a]						
	Yes	10 (83)	33 (72)	85 (81)	81 (75)	46 (85)
	No	0 (0)	7 (15)	10 (10)	15 (14)	2 (4)
	Do not know	2 (17)	6 (13)	10 (10)	12 (11)	6 (11)

[a]*P*=.055 when comparing male and female physicians.

Discussion

Principal Findings

We conducted a web-based survey study among physicians in a large hospital in Brazil to seek their opinions about the use of AI solutions in medical practice. To our knowledge, this is the first survey to interrogate physicians' expectations, fears, and opinions of AI use in medicine in Brazil. Our target population was not intended to represent the entire population of Brazilian physicians. Even so, a survey performed in a single, large private hospital can be a way of drawing attention to and starting a debate about this new subject in medicine among our physicians, as well as capturing their expectations on the topic, as AI solutions for health care workers are scarce, very limited in range, and not widespread among specialties in our reality.

Perhaps because this subject is not yet present in daily practice to most of the physicians in our study, there was a low rate of response. However, according to the HIAE marketing department, it was considered a typical respond rate since physicians usually do not respond to questionnaires in general. We wanted to extend the time frame by a couple of months for the study, but the department responsible for medical communication was concerned about overwhelming the physicians with too much electronic information. Thus, we followed the hospital's regular modus operandi.

We chose to analyze 2 different aspects: first, the number of years since graduation, divide in 3 groups: ≤10, 11-20, and >20 years; and second, sex. This was based on our personal experience and the common knowledge that younger individuals are keener to use technology applications than older individuals in general. Thus, individuals with >20 years since medical graduation would be older and might have a different approach toward AI applied to medicine than younger individuals. Additionally, female and male physicians could have different perceptions about it.

Our responders were mostly from the private sector and male and belonged to different medical areas and age groups, which was important to investigate a broader perception of medical AI. Most of the studies in medical AI acceptance were focused on specific areas, such as radiology [5-8], dermatology [1,9-11], and ophthalmology [12-14], which could be more affected than others by the adoption of AI solutions. Thus, it is important to have more general views about the topic. Our group of responders was very heterogeneous but may reflect the frequency of different specialties in the mailing list of the hospital. Thus, it would be expected to have more answers from pediatricians than psychiatrists, as an example.

We found that physicians who graduated more recently (≤10 years) showed intention to adopt AI solutions more frequently than physicians who graduated >10 years ago, likely because younger individuals have lived much of their lives using technology solutions, trusting them and relying on them for many daily tasks, and have used them more than older individuals. Furthermore, modifying old habits can be more difficult than incorporating new ones in early stages. Newly graduated medical students could already have learned about the use AI solutions in medicine, whereas older physicians would not be inclined to modify their habits as readily. Another statistically significant difference found was the more favorable opinion from female physicians about the use of AI solutions by nonphysician hospital staff. Male physicians were more conservative in that aspect, with higher number of answers against or totally against it.

Our results demonstrate that, overall, physicians have positive expectations about the use of AI in clinical practice, but they also have some concerns. Their answers indicated that medical use of AI solutions hopefully will facilitate their work and be useful for diagnoses, management, and exam interpretation. We saw a tendency for a less favorable opinion on AI for actions that are thought to be the core of physicians' responsibility. The opinion on the use of AI was 91.5% (150/164) favorable for aiding exams interpretation, 87.2% (143/164) for diagnosis, and 85.5% (140/164) for management. One-third (54/163, 33.1%) believed that the number of appointments performed by them, overall, would increase, probably by increasing the speed in making diagnosis or management decisions. Even so, they indicated that AI solutions would not interfere with their financial gain.

Probable benefits of AI solutions included greater speed, accuracy, and cost reduction for the health care system. This finding is in accordance with previous studies. A recent systematic review that included 45 studies with physicians or medical students on clinical AI showed that >60% responders had optimistic outlooks in 84% of the studies [3]. There is also an expectation that AI in medical practice will meet the higher expectations of medical treatment and physicians and will increase the efficiency of clinical care, as AI is perceived as the next big thing that will sustainably change medicine toward precision and personalized medicine [15].

Our participants believed that they would not be replaced by AI but that it would become 1 more source of information to support their work. Although the current discourse in medical literature has shifted from replacement to support of medical activities, as seen in the idea of augmented intelligence, where humans and AI work together in functions that each of them do best [16], the adoption of AI also opens the possibility of transferring decision-making to other health professionals or patients. This possibility divided their opinion, with roughly half (78/163, 47.9%) of them being against it. Many comments revealed the fear of misinterpreting the results if no medical supervision was performed. Question 18 explored the effect of AI when used by the physicians in the role of patients themselves and clearly showed that it can generate a great deal of anxiety if a troublesome diagnosis was provided by AI without specialist supervision. In 1 article focusing on patients' opinion about the use of medical AI, the patients appeared to be receptive to the use of AI for medicine if implemented in a manner that preserves the integrity of the human physician-patient relationship [1]. A review article on the convergence of human and AI poses an important statement on that matter: "Over time, marked improvements in accuracy, productivity, and workflow will likely be actualized, but whether that will be used to improve the patient-doctor relationship or facilitate its erosion remains to be seen" [17]. In China, another

study showed that the general population is more distrustful of AI in medicine, unlike the overall optimistic views posed for AI, and that the level of trust is dependent on what medical area is subject to scrutiny [19]. Those aspects are also a big concern for our physicians: worsening of the patient-physician relationship was listed right after the fear of over relying on medical AI and causing them to lose their medical skills over time. The comments showed that the benefit of human contact and the detection of emotions by the physicians cannot yet be replaced. A study with more than 1000 physicians showed that the fear of medical AI was inversely associated with advanced or intermediate AI-specific knowledge when compared with basic knowledge [6].

Possible disagreements between AI algorithms and physicians in daily practice were also explored by the questionnaire. In both questions 27 and 28, physicians believed that a third opinion should be requested (86/162, 53% and 108/160, 67.5%, respectively). Nevertheless, when the accuracy of the AI is greater the physician's (as supposed in question 28), the number of respondents who answered that the final decision should be the physician's dropped from 44.4% (72/162) to only 18.1% (29/160), revealing that the informed performance of AI solutions is crucial for physicians to make decisions.

As for legal aspects, most physicians (88/159, 55.3%) believed that the liability should be shared between them and the AI solution, reflecting the idea that the developing an AI solution involves a serious action, which requires careful engagement of all stakeholders. According to a recent article [20], all players in this field, such as physicians, developers, and health care administrators, should recognize that the implementation of an AI solution is not just a technical challenge but rather presents ethical, legal, and social challenges as well. Thus, it is important to gather all stakeholders to develop AI collaboratively from outset to implementation and evaluation [20]. It is also clear that physicians require and trust the role of government agencies to regulate this field. This is corroborated by another study [21] that discusses how regulation will become increasingly important as more algorithms start to be used in real life. Regulatory approval should not only mitigate possible harms but also define a proper balance between risks and benefits and promote effective validation standards in real settings and innovations [21]. In our study, this was more important to female physicians than male physicians.

In conclusion, our survey explored the physicians' views on AI medical solutions in a new global geographical area, showing a general positive attitude toward AI solutions, as well as some concerns, including regulation by governmental agencies, who should be using them, and the fear of physicians relying too much on them.

Limitations

It is important to note that our web-based survey was a cross-sectional study and was based on physicians' response from a single institution that has a particular interest in innovation and AI in medicine. Since the study was performed via email, physicians who answered the questionnaire could already be more likely to use technology in general. Additionally, the theme of the study was stated in the email's subject; thus, those who are interested in it would be more likely to open the email and answer the questionnaire. Therefore, the results have to be considered within a possible bias for a more positive attitude toward technology and AI in health care if compared to all physicians working in other Brazilian hospitals.

Acknowledgments

The authors thank Isabela Polizel Bonfá, who was responsible for sending the emails to the physicians.

Authors' Contributions

MG-B and BSM researched the literature and conceived the study. MG-B, BSM, and EA were involved in protocol development, gaining ethical approval, patient recruitment, and data analysis. MG-B wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire used to survey the opinion of Hospital Israelita Albert Einstein physicians on artificial intelligence.
[DOCX File, 19 KB - xmed_v5i1e50803_app1.docx]

References

1. Nelson CA, Pérez-Chada LM, Creadore A, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020 May 1;156(5):501-512. [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
2. Maassen O, Fritsch S, Palm J, et al. Future medical artificial intelligence application requirements and expectations of physicians in German University hospitals: web-based survey. J Med Internet Res 2021 Mar 5;23(3):e26646. [doi: 10.2196/26646] [Medline: 33666563]

3. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228. [doi: 10.2196/18228] [Medline: 32723713]
4. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. JAMA Netw Open 2022 Sep 1;5(9):e2233946. [doi: 10.1001/jamanetworkopen.2022.33946] [Medline: 36173632]
5. Alsharif W, Qurashi A, Toonsi F, et al. A qualitative study to explore opinions of Saudi Arabian radiologists concerning AI-based applications and their impact on the future of the radiology. BJR Open 2022 Mar 21;4(1):20210029. [doi: 10.1259/bjro.20210029] [Medline: 36105424]
6. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019 Mar 20;21(3):e12802. [doi: 10.2196/12802] [Medline: 30892270]
7. Chen M, Zhang B, Cai Z, et al. Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. Front Med (Lausanne) 2022 Aug 31;9:990604. [doi: 10.3389/fmed.2022.990604] [Medline: 36117979]
8. Sharma A, Minh Duc NT, Luu Lam Thang T, et al. A consensus-based Checklist for Reporting of Survey Studies (CROSS). J Gen Intern Med 2021 Oct;36(10):3179-3187. [doi: 10.1007/s11606-021-06737-1] [Medline: 33886027]
9. Huisman M, Ranschaert E, Parker W, et al. An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education. Eur Radiol 2021 Nov;31(11):8797-8806. [doi: 10.1007/s00330-021-07782-4] [Medline: 33974148]
10. Huisman M, Ranschaert E, Parker W, et al. An international survey on AI in radiology in 1041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude. Eur Radiol 2021 Sep;31(9):7058-7066. [doi: 10.1007/s00330-021-07781-5] [Medline: 33744991]
11. Ooi SKG, Makmur A, Soon AYQ, et al. Attitudes toward artificial intelligence in radiology with learner needs assessment within radiology residency programmes: a national multi-programme survey. Singapore Med J 2021 Mar;62(3):126-134. [doi: 10.11622/smedj.2019141] [Medline: 31680181]
12. Polesie S, Gillstedt M, Kittler H, et al. Attitudes towards artificial intelligence within dermatology: an international online survey. Br J Dermatol 2020 Jul;183(1):159-161. [doi: 10.1111/bjd.18875] [Medline: 31953854]
13. Pangti R, Gupta S, Gupta P, Dixit A, Sati HC, Gupta S. Acceptability of artificial intelligence among Indian dermatologists. Indian J Dermatol Venereol Leprol 2022;88(2):232-234. [doi: 10.25259/IJDVL_210_2021] [Medline: 35170270]
14. Cho SI, Han B, Hur K, Mun JH. Perceptions and attitudes of medical students regarding artificial intelligence in dermatology. J Eur Acad Dermatol Venereol 2021 Jan;35(1):e72-e73. [doi: 10.1111/jdv.16812] [Medline: 32852856]
15. Scheetz J, Rothschild P, McGuinness M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 2021 Mar 4;11(1):5193. [doi: 10.1038/s41598-021-84698-5] [Medline: 33664367]
16. Zheng B, Wu MN, Zhu SJ, et al. Attitudes of medical workers in China toward artificial intelligence in ophthalmology: a comparative survey. BMC Health Serv Res 2021 Oct 9;21(1):1067. [doi: 10.1186/s12913-021-07044-5] [Medline: 34627239]
17. Valikodath NG, Al-Khaled T, Cole E, et al. Evaluation of pediatric ophthalmologists' perspectives of artificial intelligence in ophthalmology. J AAPOS 2021 Jun;25(3):164. [doi: 10.1016/j.jaapos.2021.01.011] [Medline: 34087473]
18. SurveyMonkey. URL: https://www.surveymonkey.com/ [accessed 2024-03-12]
19. Ye T, Xue J, He M, et al. Psychosocial factors affecting artificial intelligence adoption in health care in China: cross-sectional study. J Med Internet Res 2019 Oct 17;21(10):e14316. [doi: 10.2196/14316] [Medline: 31625950]
20. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast 2020 Feb;49:25-32. [doi: 10.1016/j.breast.2019.10.001] [Medline: 31677530]
21. Allen TC. Regulating artificial intelligence for a successful pathology future. Arch Pathol Lab Med 2019 Oct;143(10):1175-1179. [doi: 10.5858/arpa.2019-0229-ED] [Medline: 31268373]

Abbreviations

AI: artificial intelligence
CROSS: Checklist for Reporting Survey Studies
HIAE: Hospital Israelita Albert Einstein
ICF: informed consent form

Edited by E Meinert; submitted 12.07.23; peer-reviewed by F Baglivo, L Carrasco-Ribelles; revised version received 28.12.23; accepted 13.01.24; published 25.03.24.

Please cite as:
Giavina-Bianchi M, Amaro Jr E, Machado BS
Medical Expectations of Physicians on AI Solutions in Daily Practice: Cross-Sectional Survey Study
JMIRx Med 2024;5:e50803
URL: https://xmed.jmir.org/2024/1/e50803
doi:10.2196/50803

Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers

David Bychkov, PhD

UC Institute for Prediction Technology, University of California, Irvine, Orange, CA, United States

Corresponding Author:
David Bychkov, PhD

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.08.14.23294073v1

Companion article: https://med.jmirx.org/2024/1/e57701

Companion article: https://med.jmirx.org/2024/1/e57159

Companion article: https://med.jmirx.org/2024/1/e57116

Abstract

Background: To address the pandemic, the Defense Health Agency (DHA) expanded its TRICARE civilian provider network by 30.1%. In 2022, the DHA Annual Report stated that TRICARE's provider directories were only 80% accurate. Unlike Medicare, the DHA does not publicly reveal National Provider Identification (NPI) numbers. As a result, TRICARE's 9.6 million beneficiaries lack the means to verify their doctor's credentials. Since 2013, the Department of Health and Human Services' (HHS) Office of Inspector General (OIG) has excluded 17,706 physicians and other providers from federal health programs due to billing fraud, neglect, drug-related convictions, and other offenses. These providers and their NPIs are included on the OIG's List of Excluded Individuals and Entities (LEIE). Patients who receive care from excluded providers face higher risks of hospitalization and mortality.

Objective: We sought to assess the extent to which TRICARE screens health care provider names on their referral website against criminal databases.

Methods: Between January 1-31, 2023, we used TRICARE West's provider directory to search for all providers within a 5-mile radius of 798 zip codes (38 per state, ≥10,000 residents each, randomly entered). We then copied and pasted all directory results' first and last names, business names, addresses, phone numbers, fax numbers, degree types, practice specialties, and active or closed statuses into a CSV file. We cross-referenced the search results against US and state databases for medical and criminal misconduct, including the OIG-LEIE and General Services Administration's (GSA) SAM.gov exclusion lists, the HHS Office of Civil Rights Health Insurance Portability and Accountability Act (HIPAA) breach reports, 15 available state Medicaid exclusion lists (state), the International Trade Administration's Consolidated Screening List (CSL), 3 Food and Drug Administration (FDA) debarment lists, the Federal Bureau of Investigation's (FBI) list of January 6 federal defendants, and the OIG-HHS list of fugitives (FUG).

Results: Our provider search yielded 111,619 raw results; 54 zip codes contained no data. After removing 72,156 (64.65%) duplicate entries, closed offices, and non-TRICARE West locations, we identified 39,463 active provider names. Within this baseline sample group, there were 2398 (6.08%) total matches against all exclusion and sanction databases, including 2197 on the OIG-LEIE, 2311 on the GSA-SAM.gov list, 2 on the HIPAA list, 54 on the state Medicaid exclusion lists, 69 on the CSL, 3 on the FDA lists, 53 on the FBI list, and 10 on the FUG.

Conclusions: TRICARE's civilian provider roster merits further scrutiny by law enforcement. Following the National Institute of Standards and Technology 800, the DHA can mitigate privacy, safety, and security clearance threats by implementing an insider threat management model, robust enforcement of the False Claims Act, and mandatory security risk assessments. These are the views of the author, not the Department of Defense or the US government.

(JMIRx Med 2024;5:e52198) doi:10.2196/52198

KEYWORDS

TRICARE; health care fraud; Defense Health Agency; fraud; fraudulent; insurance; coverage; beneficiary; beneficiaries; background check; background checks; demographic; security clearance; FDA; Medicaid; Medicare; provider; provider referral; military; false claims act; HIPAA breach; OIG-LEIE; inspector general; misconduct; insider threat; information system; zero trust; data management; Food and Drug Administration; Health Insurance Portability and Accountability Act breach; Office of the Inspector General's List of Excluded Individuals and Entities

Introduction

In response to the COVID-19 pandemic, the Military Health System expanded access to civilian providers. From 2020 to 2021, TRICARE, the Defense Health Agency (DHA) insurance scheme, maintained a steady beneficiary pool of 9.6 million military beneficiaries. Meanwhile, its civilian roster ballooned from 548,297 to 713,395 providers, a 30.11% increase over 1 fiscal year [1]. According to a report by the Department of Defense's Office of the Inspector General (DODIG), the DHA's Program Integrity (PI) office suspended medical record audits related to improper payments during the pandemic due to a lack of in-person investigators [2]. In 2021, the DHA-PI received 600 lead requests, opened 110 new cases, and managed 693 active cases. According to their annual reports, they sanctioned no additional health care providers since August 2020 [1,3]. Furthermore, only 80% of the provider directory information published by TRICARE's managed care support contractors (MCSCs) was accurate [1]. Due to call center problems and technical challenges with the Defense Enrollment Eligibility Reporting System, the percentage of health care provider contracts compliant with TRICARE fluctuated between 79.5% and 94.1% throughout the first 47 months of the current T2017 contract [1]. Nevertheless, TRICARE promises beneficiaries that their affiliated civilian doctors meet "stringent quality and credentialing requirements" [1]. In the United States, National Provider Identification (NPI) numbers are the sole identifier for licensed clinicians. TRICARE does not publish NPIs in its provider directory. Military treatment facilities (MTFs), service members, and their families, therefore, lack a simple way to verify the bona fides of outside civilian providers.

Health care organizations may not purchase goods or services from excluded entities and vendors without jeopardizing their federal contracts [4,5]. The Centers for Medicare and Medicaid (CMS) require federally funded health care organizations to screen out providers against two sources at regular intervals: the Office of Inspector General's (OIG) List of Excluded Individuals and Entities (LEIE) and the General Services Administration's (GSA) SAM.gov exclusion list [6]. The OIG-LEIE is a comprehensive registry that excludes individuals and entities from participating in federally funded health care programs for a range of reasons, including patient abuse or neglect, billing fraud, and drug-related convictions [7]. According to Burton et al's [8] demographic analysis of 1289 physicians on the OIG-LEIE between January 2008 and December 2013, a total of 509 were excluded for license revocation or suspension (Social Security Act 1128 (b)(4)), 280 for Medicare/Medicaid fraud conviction (Social Security Act 1128 (a)(1)), 123 for another type of health fraud conviction (Social Security Act 1128 (a)(1)), and 191 for felony controlled substance disbursement (Social Security Act 1128 (a)(4)) [8]. Male physicians represented nearly 85% of the total excluded physicians but accounted for nearly 70% of the general physician population [8]. One long-term care facility was fined US $376,000 for multiple violations of exclusion rules, with fines typically exceeding tens of thousands of dollars per violation [9]. In FY2022, the Department of Health and Human Services' (HHS) OIG Medicaid Fraud Control Unit reported 10,604 and 7202 open criminal and civil investigations, respectively [10].

The GSA-SAM.gov is a list of corporations forbidden from doing business with the US government. In April 2003, the HHS debarred a medical supplies company for 5 years after the owner pleaded guilty to Medicare fraud [11]. As HHS did not debar the individual's company, he transferred ownership of the company to his wife and they continued the scheme. After investigators discovered the corporate change, the couple transferred the company ownership again to a neighbor. Two years later, the neighbor sold the company back to the original owner's wife. To prevent further discovery, the wife changed her last name to her maiden name. Thanks to these tactics, the couple managed to defraud federal health programs for the entire 5 years of debarment. Under procedures outlined in Federal Acquisition Regulation (FAR) subpart 9.4.19, an agency suspending and debarring official may suspend any contractor upon receiving an allegation that a contractor is not acting responsibly [12]. The suspension is enacted by listing the contractor in the excluded status on the GSA-SAM.gov list and notifying the contractor in writing. Under 48 C.F.R. § 9.405, no award can be issued to a contractor suspended, proposed for debarment, debarred, or otherwise award ineligible unless the agency head or designee determines in writing that a "compelling reason" exists [5]. For the DHA to conduct business with an excluded health provider, as per 10 U.S.C. § 2393, the Secretary of Defense must provide the GSA notice of the "Compelling Reason Determination Pursuant to 48 C.F.R. § 9.405" for publication on the web [13]. Thirty agencies contribute data to the Interagency Suspension and Debarment Committee (ISDC) and the GSA-SAM.gov list. As of the time of publication, the ISDC has published no compelling reason determinations associated with TRICARE's civilian provider network [14].

Direct or indirect federal reimbursement for goods or services rendered by an excluded individual or entity is prohibited by the False Claims Act, FAR 9.404 "Exclusions in the System for Award Management'' and the Civil Monetary Penalties Law [15]. This includes reimbursement for salaries, benefits, or items claimed or billed by licensed health care providers and administrative personnel. Hospitals, equipment suppliers, drug manufacturers, and health management organizations that serve federal programs must use the OIG-LEIE and GSA-SAM.gov to screen out inappropriate employees and contractors [16]. Billing federal health care programs for services rendered by

excluded providers can result in a minimum penalty of US $10,000 per instance [17]. To automate this process, McKesson, a revenue cycle management and electronic health record software provider has integrated exclusion monitoring tools into their products [18]. As TRICARE still accepts billing claims by fax and mail, fraudsters can potentially thwart automated exclusion screening processes [19].

The DHA provides no mandatory information security training to outside contractors. Although the Health Insurance Portability and Accountability Act (HIPAA) of 1996 requires health care providers to perform security risk assessments (SRAs) [20], 17% of respondents to the 2021 Healthcare Information and Management Systems Society Healthcare Cybersecurity Survey reported not having a budget for risk assessments [21]. In the same survey, 83% of respondents had experienced a cyberattack. Due to budget or logistical concerns, 26% had reduced their overall cybersecurity budget [21]. Continuous personnel screening is an effective mitigation practice against insider threats. Malicious nonstate actors seek negligent insiders to help them target personal identifying information (PII) and personal health information (PHI) contained in electronic health records at medical practices [22].

Approximately 18% of all service members receive security clearances [23]. Service members need to discuss their medical conditions with health providers without fear of data breaches or blackmail by an adversary [24]. Military personnel have a reasonable expectation that TRICARE-credentialed health care providers are not fugitives from justice, in violation of international sanctions, a threat to national security, or associated with a cyber breach. Unfortunately, military-affiliated consumers are 76% more likely than other adults to experience medical benefit fraud and identity theft [25]. Nicholas et al [26] performed a cross-sectional study of 8204 Medicare beneficiaries who received care from excluded providers. They revealed that patients treated by fraudsters experienced a 13%-23% increased risk of mortality and an 11%-30% higher risk of hospitalization.

In addition to the OIG-LEIE and GSA-SAM.gov, multiple public databases exist to search names concerning each of these issues, including:

- HHS' Office of Civil Rights' HIPAA Breach Report affecting 500 or more patients (HIPAA) [27];
- HHS-OIG list of fugitives (FUG) wanted for health care fraud, abuse, or child support obligations [28];
- International Trade Administration's Consolidated Screening List (CSL) of parties for which the US government maintains restrictions on exports, reexports, or transfers of items [29];
- The Federal Bureau of Investigation's (FBI) list of January 6th capitol breach defendants [22,30];
- Lists of providers excluded by 15 of 21 state Medicaid programs for fraud, neglect, and abuse (state) [31];
- The US Food and Drug Administration's (FDA) debarment lists for illegal drug imports, food imports, and drug product activity [32]; and
- DHA's Sanction List for TRICARE-specific billing fraud and patient abuse (DHA sanctions) [33].

This study, therefore, aims to determine if TRICARE refers its beneficiaries to providers found on government exclusion or sanction lists. If so, we aim to identify their professional and geographic characteristics. Finally, we offer recommendations to mitigate potential threats posed by excluded providers to the safety, privacy, and security clearances of service members.

Methods

Overview of the Study Area

Time constraints imposed by the 1-year time limit of our Human Research Protection Program required us to choose and investigate only one MCSC's provider roster. TRICARE's two main MCSCs, Health Net Federal Services, LLC (TRICARE West) and Humana Inc (TRICARE East) [1], receive approximately US $7.2 billion and US $7.87 billion per year, respectively, to ensure military medical readiness in their respective regions [34]. Additional MCSCs provide services to niche beneficiary populations, including Johns Hopkins Medicine (US Family Health Plan) and International SOS Government Services, Inc (TRICARE Overseas) [1]. These MCSCs do not display the NPIs of their health care providers on the web. From this group, we chose TRICARE West randomly.

TRICARE West is currently operated by Health Net Federal Services, LLC across 21 states: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Iowa (excluding the Rock Island Arsenal area), Kansas, Minnesota, Missouri (excluding the St. Louis area), Montana, Nebraska, Nevada, New Mexico, North Dakota, Oregon, South Dakota, Texas (Amarillo, Lubbock, and El Paso areas only), Utah, Washington, and Wyoming.

Providers

To gather further information on health care providers in the search area, we evaluated national and statewide trends from the Health Research Services Administration's (HRSA) National Practitioner Data Bank related to licensure, adverse events, malpractice, Drug Enforcement Administration enforcement, and exclusions [35]. HRSA also publishes zip code–level data on health provider shortage areas (HPSAs) [36] and the availability of nearby Federally Qualified Health Centers (FQHCs). FQHCs are nonprofit medical facilities that provide primary care to an area or people in need, offer a sliding fee scale, provide complete services, have a quality assurance program, and maintain a governing board of directors [37]. For rural communities where TRICARE providers refuse new patients or lack available appointment times, FQHCs may be an excellent closer option than an MTF. They compare favorably to private providers for patient access, safety, and satisfaction [38]. Due to their nonprofit mission, however, many FQHCs lack digital health care IT resources [39]. No studies have investigated TRICARE beneficiary use of FQHCs.

The Patient Population

TRICARE West's 3.7 million patients reside primarily in Texas, California, Washington, Colorado, and Arizona [40]. The Defense Manpower Data Center provides some details on the distribution of military service members and their families. For example, Texas is home to the highest concentration of service

members from both the Army (16.9%) and Air Force (12.9%). California houses the largest number of active duty Marines (36.6%) [41]. Most active Space Force Guardians (24.7%), on the other hand, are based in Colorado [41]. By contrast, the largest concentration of Navy active duty members reside in Virginia (26.5%) [42]. Although one-third of service members move every year, no public-facing data indicate how many TRICARE West beneficiaries transition to TRICARE East or vice versa [43]. While service members are somewhat healthier than civilians, today's TRICARE providers must treat the same conditions that impact the rest of society, such as cardio- and neurovascular diseases, sexually transmitted infections, substance use disorders, metabolic disorders, and mental health issues [44-46]. According to the Medical Expenditure Panel Survey, a nationwide questionnaire, TRICARE-covered families nationwide reported inferior access to medical care when compared to uninsured, commercially insured, and privately insured peers [47]. Furthermore, military families dealing with complex pediatric care reported worse outcomes than civilians [47]. Approximately 40% of military families have children [48]. At least 1 in 12 rely on Medicaid to provide supplemental coverage for their children [49].

Data Sources

TRICARE West publishes the first and last names, specialty type, practice type, company names, and contact information of TRICARE-credentialed civilian providers on their public-facing provider directory, which was accessed on the TRICARE West website [50] for this study.

Between January 1-31, 2023, we used TRICARE West's provider directory to search for all names within a 5-mile radius of 798 United States Postal Service–designated zip codes within TRICARE West's territory of 12,574 zip codes. We copied and pasted all raw results into an Excel (.xls) file (Microsoft Corporation).

To ensure compliance with the terms and conditions of the TRICARE provider directory, all data were manually accessed.

Zip Code Search Selection

To ensure each batch of results included the largest possible number of TRICARE provider names, we limited the scope of each search query on the TRICARE West provider directory to those zip codes known to contain at least 10,000 residents and 1 credentialed medical provider. Population estimates were gathered from the 2020 US Census [51]. For each state, 38 zip codes were selected for searching on the TRICARE West provider directory. These 798 zip code searches represent 6.1% of TRICARE West's total land coverage area.

Provider Search Parameters

The TRICARE West provider directory displays first, last, and business names; full addresses; phone and fax numbers; degree type; provider gender; specialty; and active/closed status.

Analysis Parameters

To evaluate the data, we gathered the most current .xls versions of the OIG-LEIE and GSA-SAM.gov, the HIPAA list, state lists, CSL, FDA lists, the FBI list, the DHA Sanctions List, and FUG. Using the VLOOKUP function in Excel, we cross-referenced the providers' first, last, and corporate names (with and without zip code qualifiers) against each exclusion, sanction, and violation list. VLOOKUP search strings are useful database tools for detecting fraud patterns in spreadsheets, including common names or locations [52,53].

Ethical Considerations

This study relied on no confidential data. It was conducted with an exemption from the Human Resource Protection Program of Defense Acquisition University, received on January 30, 2023.

Results

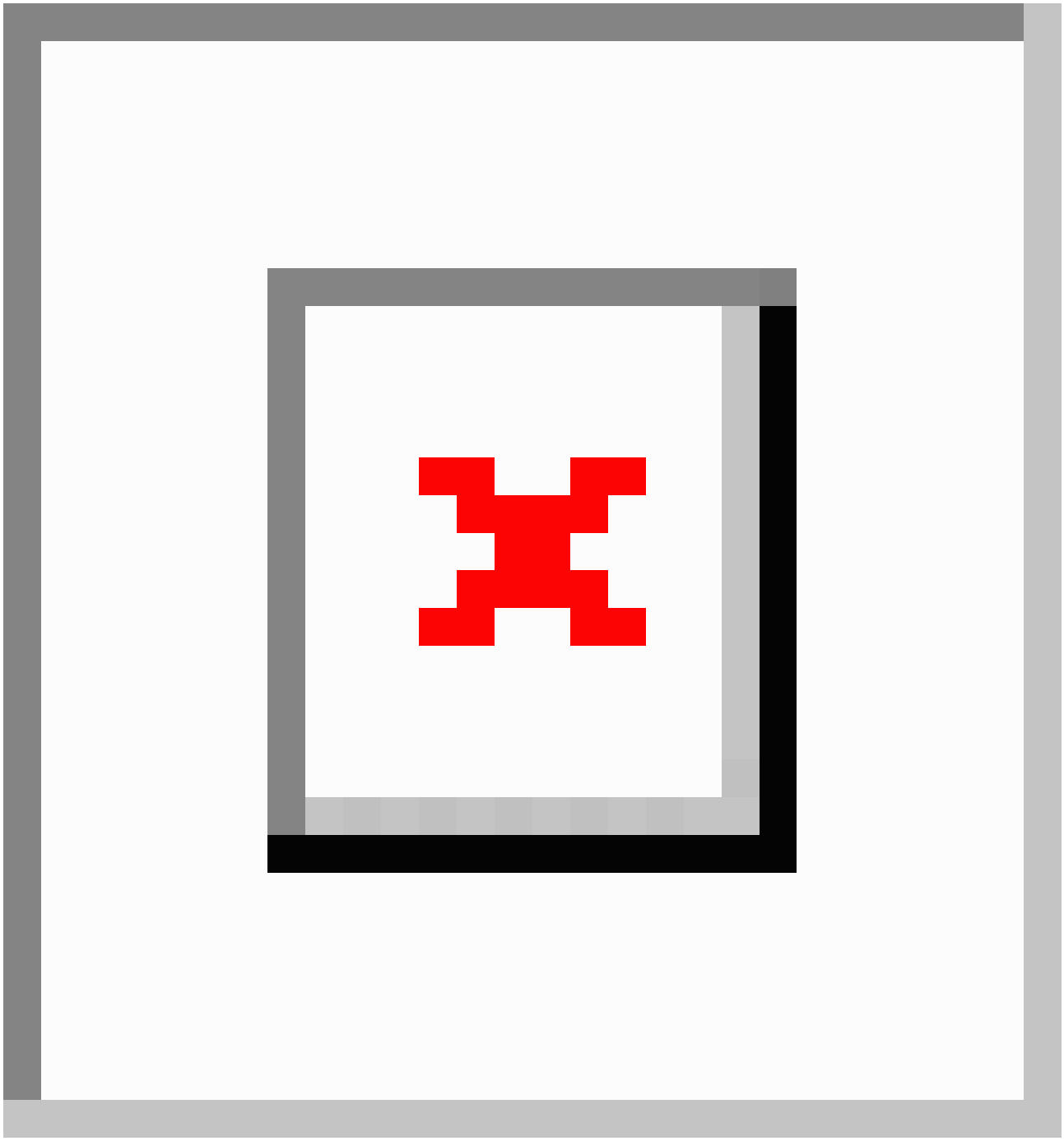
Overview

Our search of the TRICARE West provider directory yielded 111,619 raw results across the 798 zip code areas. Searches of 54 zip codes yielded no entries, including 21 searches in Missouri, 18 in Wyoming, 5 in Iowa, 5 in Alaska, 4 in Washington, and 1 in Texas.

After filtering out 72,156 (64.65%) entries for duplicates, closed offices, and data located in external states, we established our baseline list of 39,463 TRICARE West active provider names. This group accounts for 5.53% of TRICARE's 2021-2022 nationwide civilian roster [1].

Within the baseline group, 2398 (6.08%) provider names matched the first and last names of individuals and business owners found on 10 federal and state regulatory watch lists (Figure 1).

Figure 1. Map of search area zip codes (in blue) vs zip codes where provider names were matched against exclusion lists (in red). Map data ©2024 Google, INEGI.



Exclusion Types

Among the 2398 names, 2197 appear on the OIG-LEIE and 2311 appear on the GSA-SAM.gov. Within the group, 2 appear on the HIPAA list, 69 appear on the CSL list, 10 appear on the FUG, 15 appear on the FDA lists, 53 appear on the FBI list, and 54 appear on 15 different state Medicaid exclusion lists (Alaska: 0; California: 38; Hawaii: 0; Idaho: 0; Iowa: 1; Kansas: 0; Minnesota: 0; Missouri: 0; Montana: 0; North Dakota: 0; Nebraska: 0; Nevada: 1; Texas: 20; Washington: 0; Wyoming: 0).

Our search matched 1997 providers with 2 total exclusion types, 230 providers with 1 type, 158 providers with 3 exclusion types, 12 providers with 4 exclusion types, and 1 provider with 5 exclusion types. All providers with 2 or more exclusions have their first, last, or corporate names appear on the OIG-LEIE and GSA-SAM.gov exclusion lists. All names that appear on the FDA lists, HIPAA list, FUG, and CSL also appear on the OIG-LEIE or GSA-SAM.gov. Most providers with names on the FBI list (50/54) appear on the GSA-SAM.gov.

One name matched 1 provider on the DHA's Sanctioned Provider List for TRICARE-specific offenses. To protect the provider's privacy, we will not identify their state or other matching characteristics. Since 1990, the DHA Sanctions List has included no new providers in Utah or Minnesota (Table 1).

Table . Exclusion search vs actual exclusion enforcement by TRICARE.[a]

Top 5 states for exclusion-provider name matches	State TRICARE population, n	Historical enforcement by TRICARE, n	Exclusions found for TRICARE West providers, n
Utah	80,390	0	264
Minnesota	72,931	0	227
Kansas	120,503	6	212
Colorado	253,214	1	201
Washington	348,694	3	180
Total	N/A[b]	10	1084

[a]Since 1990, the Defense Health Agency has sanctioned 6 providers in Kansas, 1 in Colorado, and 3 in Washington State. They sanctioned no providers in Utah or Minnesota, the states where we found the highest concentration of names with exclusions.

[b]N/A: not applicable.

Provider Characteristics

Provider names linked to exclusions were also associated with a medical degree (doctor of medicine [MD] n=1288, 54%; doctor of osteopathic medicine [DO] n=199, 8%). Family medicine was the top specialty (n=324, 13.5%), followed by nurse practitioner (n=148, 6.2%), internal medicine (n=112, 4.7%), optometrist (n=99, 4.1%), and pediatrics (n=91, 3.8%). Diploma types with the fewest exclusions included master of nursing (MN; n=1), registered behavior technician (RBT; n=1), licensed clinical psychologist (LCP; n=1), physical therapist (PT; n=1), and certified registered nurse anesthetist (CRNA; n=1). Our results include information about provider specialty and diploma type. Our results also included information on provider gender. Male providers accounted for 59.42% of the exclusions, while female providers accounted for 40%. Two providers did not report their gender.

Provider Location

States with the highest concentrations of names associated with exclusions within the TRICARE West network were Utah (11%) and Minnesota (9%). The five zip codes with the highest exclusions were 84096 (Herriman, UT; n=18), 84062 (Pleasant Grove, UT; n=17), 99669 (Soldotna, AK; n=16), 84790 (Washington County, UT; n=14), and 80524 (Fort Collins, CO; n=14). The five zip codes with only 1 total exclusion were 96782 (Honolulu, HI; n=1), 79159 (Amarillo, TX; n=1), 84003 (Utah County, UT; n=1), 83401 (Bonneville County, ID; n=1), and 99505 (Anchorage, AK; n=1).

Finally, we conducted a follow-up TRICARE West directory search of how many MTFs operate within a 100-mile radius of those zip codes with the highest concentration of excluded provider name matches (Table 2). Whereas 96782 (Honolulu, HI) had 9 MTFs within a 100-mile radius and 1 provider associated with an exclusion, 83401 (Bonneville County, ID), 84790 (Washington County, UT), and 79159 (Amarillo, TX) have no MTF alternatives to provider names associated with exclusions.

Table . Top 10 zip codes for provider names matched to exclusions.

Top 10 zip codes	Total names, n	MTFs[a] within 100 miles, n
84096 (Herriman, UT)	18	2
84062 (Pleasant Grove, UT) [b]	17	2
99669 (Soldotna, AK)	16	2
84790 (Washington County, UT)	14	0
80524 (Fort Collins, CO)	14	2
96782 (Honolulu, HI)	1	9
79159 (Amarillo, TX)	1	0
84003 (Utah County, UT)	1	2
83401 (Bonneville County, ID)	1	0
99505 (Anchorage, AK)	1	2

[a]MTF: military treatment facility.

[b]Italicized zip codes are federally designated health provider shortage areas.

Discussion

Significance of Findings

This is the first academic study to assess the level of compliance of a federal health care program's provider directory with criminal background check laws. In addition to evaluating provider names on databases for medical fraud, we expanded our search to include domestic terrorism, financial crimes, child support delinquency, and data breaches. Our matches included 28 fugitives from justice and 58 January 6th defendants.

Our results align with historical trends contained in Total Force Medical Readiness (TFMR) reports [54]. Between 2013 and 2021, overall individual medical readiness among nondeployed military components dropped 6 points [1]. In Q4 2013, 75% of the total force reported being "fully medically ready" (ie, satisfactory dental health, completion of periodic health assessments, deployment-limiting medical conditions status, current immunization status, completion of medical readiness lab tests, and possession of required individual medical equipment). By Q4 FY2021, readiness dropped to 69% [1]. Currently, TFMR reports do not indicate if they rely on data providers who received improper payments. Furthermore, they display no margin of error.

Our demographic findings are consistent with recent and historical reviews of excluded providers. In a cross-sectional study on physician exclusions from 2007 to 2017, Chen et al [55] reported that the total number of physician exclusions grew by 20% to include nearly 0.3% of all US physicians. Exclusions are more common in the West and Southeast among male physicians. In line with Burton et al's [8] demographic analysis of the OIG-LEIE the majority of the provider names we sampled from TRICARE West's provider directory have specialty training in family medicine and appear on the GSA-SAM.gov.

Geospatial data may help investigators link medical fraud and adverse events. According to HRSA's database of HPSAs, 8 of the top 10 zip codes identified by our study for provider names with exclusions lack primary care workers [36]. Indeed, Zhang et al [56] forecasted that Western states will face an acute shortage of 69 physicians per 100,000 residents by 2030.

During the pandemic, HRSAs reported an increase in the number of adverse action reports and medical malpractice payments within the TRICARE West states with the most provider name-exclusion matches (Table 1). The most common adverse event is a patient fall [57]. Increased patient falls at hospitals and nursing homes are typically caused by breakdowns in clinical communications, including systemwide failures in teamwork and failures to consistently follow policies [58]. In Utah, for example, there were 359 adverse event reports in 2019—the highest number recorded in the state's history [59]. In 2013, there were 555 adverse events in Colorado. By 2020, they peaked at 1215 [57].

We matched the most provider names against exclusions in states where the DHA sanctioned few or no physicians. For example, we detected the highest total number of potentially excluded providers in Utah (n=264), Minnesota (n=227), Kansas (n=212), Colorado (n=201), and Washington (n=180; Table 1).

We found the lowest number of exclusion-provider name matches in Idaho (n=32), Missouri (n=21), Nebraska (n=15), and Montana (n=14). The TRICARE West states with the lowest number of DHA-sanctioned providers since 1990 were Utah (n=0), Minnesota (n=0), Colorado (n=1), Oregan (n=0), Montana (n=1), North Dakota (n=0), South Dakota (n=0), and Washington (n=3). Whereas the OIG-LEIE contains a total of 77,621 providers banned from federal health programs, the DHA Sanctions List contains only 129 provider names. In other words, medical fraudsters may have evaded detection by operating in states where the DHA only checks provider names against their own sanctions list.

TRICARE's financial statement reveals opportunities for increased enforcement of the False Claims Act. The 2022 DHA Annual Report shows US $900,000 collected from excluded providers in 2020 and US $100,000 in 2019—down from a high of US $1.4 million in 2017 [1]. In their FY2021 evaluation of the TRICARE Program, the DHA reported an average annual expenditure of US $2251 on the typical single beneficiary [1]. If 120 (5%) of the 2398 identified TRICARE West providers billed for services for 3 beneficiaries over the preceding calendar year, the DHA could recoup at least US $810,360. Under the False Claims Act, the agency could require each of those providers to pay an additional US $10,000 per patient in penalties (ie, US $3,600,000 in penalties; US $4.4 million in total improper payment recoveries) [60].

Recommendations

Service members and their families need tools to protect themselves from identity theft and medical fraud. HIPAA requires health care providers, including TRICARE West, to safeguard PHI [61]. HIPAA requires health care organizations to conduct regular SRAs to evaluate external and insider threats to PHI and other sensitive data. In addition to their appearance on the OIG-LEIE, two provider names on TRICARE West's roster were associated with data breaches impacting over 500 medical patients. At least 450,000 TRICARE beneficiaries have an active security clearance [62]. TRICARE has not published a data breach disclosure since 4.9 million patient records disappeared in 2011 [63].

All federal agencies must implement a zero trust (ZT) architecture by 2027. The zero trust paradigm of information security requires participants to operate on a "need to know" basis. We recommend that the DHA prioritize the deployment of a ZT-based Insider Threat Management Model to protect the PHI of service members with security clearance (Table 3). Our model continuously screens the DHA's civilian providers against federal and state databases. Aligned with the National Institute of Standards and Technology (NIST) 800-207 [64] and the CISA Zero Trust Maturity Model [65], it helps TRICARE administrators wall off providers linked to exclusions from beneficiaries and their data. Effectively, TRICARE patients with confidential-level security clearance or higher gain access to a filtered roster of health care providers. If a search of TRICARE's website by a National Security Agency engineer with a Top Secret or Sensitive Compartmented Information clearance in Maryland for "John + Smith" matches the name of an excluded provider on the OIG-LEIE and GSA-SAM.gov

in a beneficiary's state, the roster would only display licensed, bona fide physicians. If none exist nearby, the roster would display the closest MTF. If no MTFs exist within a 100-mile radius, TRICARE's website could display nonemergency medical transportation options to the nearest MTF, secure telehealth options, and covered care options at the nearest out-of-network screened provider. According to our Insider Threat Management Model, any providers in the TRICARE network associated with 3 or more exclusions would lose all access to *any* beneficiaries with clearance. Those beneficiaries need to be transferred to the nearest MTF. Patients who receive care from providers with 3 or more exclusions face higher risks of adverse events [9].

Table . Insider Threat Management Model. The system continuously vets all civilian network providers.

Beneficiary security clearance level	Doctor's exclusions		
	1 exclusion	2 exclusions	≥3 exclusions
Confidential	Yellow	Yellow	Red
Secret	Yellow	Orange	Red
Top Secret	Yellow	Red	Red
Sensitive	Yellow	Red	Red
Sensitive Compartmented Information	Yellow	Red	Red

HIPAA entitles TRICARE beneficiaries to specific tools for medical privacy. For example, they may opt out of the Joint Health Information Exchange (JHIE), an electronic platform for transmitting medical data to civilian providers [7]. Currently, the JHIE opt-out system requires a paper-pencil request. As an alternative, the DHA could offer an electronic JHIE opt-out button in current Defense Finance and Accounting Service (DFAS) myPAY dashboards. DFAS myPAY offers multifactor authentication and paperless transactions to mitigate the threat of lateral nonauthorized movement of PII and PHI beyond the control of the beneficiary [66,67]. Our Insider Threat Management Model complements rather than replaces current federal agency mandates for periodic security training of all DHA employees, contractors, and credentialed providers; implementation of strict password, SRAs, and account management practices; explicit security agreements and access restrictions; PHI, PII, and other sensitive information only made available to those who require it; and use of a security information and event management [68] solution. Our model aligns with the CERT Resilience Management Model [69], ISO/IEC 27002:2022 [70], the NIST Privacy Framework [71], and the NIST Cybersecurity Framework [72]. Although TRICARE operation manuals provide clear guidance to providers related to background checks, patient privacy, and cyber hygiene [73], no single clearinghouse provides ethics training, employee background checks, security information and event management solutions, and SRAs for their civilian providers.

By publishing NPI numbers alongside provider contact details, the DHA could reduce the likelihood of fraudulent claims and improper payments. Whereas NPI numbers are permanent, providers may change or add business names, last names, locations, specialties, and state registrations. Furthermore, no free web-based search tool continuously gathers data from all state, federal, and licensing board databases using a combination of name and address spellings.

Limitations

Our study was conducted under strict resource and time constraints. Between June 2022 and June 2023, we filed 4 Freedom of Information Act (FOIA) requests for (1) NPI numbers of TRICARE civilian providers, (2) data on TRICARE provider roster use, (3) data on HIE Opt-Out requests, and (4) data related to HIPAA-mandated SRAs (also known as "Compliance Risk Assessments") performed by MCSCs and TRICARE network providers. Although the DHA acknowledged each request, they fulfilled none as of the date of this publication. The DHA's most recent FOIA disclosure was on July 8, 2015 [74].

This study identified two TRICARE West provider names with matches on the HIPAA breach list. The HHS Office of Civil Rights does not require medical practices to report privacy breaches that impact ≤499 patients. The vast majority of civilian health providers operate in small practices and do not conduct regular SRAs [21].

The DHA's annual report states that 80% of the public-facing information of their provider directory is accurate [7]. It does not, however, indicate which portion of the provider directory is not accurate. The TRICARE West provider does not include middle initials with all names and a complete list of states in which their providers are licensed. To address these limitations, we spent time manually differentiating providers with common names. For example, our list of 2398 matches includes a common female name that appears 10 times. Each of these female providers lives in different zip codes or states, has different degrees, and practices completely unrelated types of medicine. In other words, expert review was necessary to ensure our final sample was valid and free of duplicates.

In April 2023, the DHA announced that they would replace Health Net Federal Services with TriWest Healthcare Alliance as the prime contractor for TRICARE West [75]. TriWest has made no announcements regarding the transition of TRICARE West's provider referral website.

Conclusion

Our study reveals that 6.08% of the provider names listed on TRICARE West's provider directory match individuals listed on federal and state exclusion lists. To assist law enforcement, we provided all data and study materials to the DODIG on May 8, 2023, and the DHA-OIG on May 24, 2023, in the form of a whistleblower complaint. To triage future threats associated with excluded and sanctioned provider names, we proposed a zero trust–based Insider Threat Management Model for TRICARE beneficiaries with security clearances. In future studies, we intend to compare TRICARE East, Medicare, the Children's Health Insurance Program, and Substance Abuse and Mental Health Services Administration provider rosters against a broader spectrum of exclusion, sanction, and violation categories (ie, the Federal Sex Offender Registry). We also intend to develop products and interventions to automate background checks, protect patient privacy, and educate health care administrators about insider threats.

Acknowledgments

The author would like to express gratitude to his former colleagues on the Cyber Enterprise Team at Defense Acquisition University. Dr Paul Shaw provided substantial feedback and encouragement at each stage. Dr Aaron Jacobson provided feedback on the Insider Threat Management Model and reviewed data. The author would also like to thank the Department of Defense Inspector General for their willingness to review the data.

The views expressed in this article are those of the author. They do not reflect the official policy or position of the Department of Defense, Defense Acquisition University, or the US Government. This work was prepared as part of the author's official duties as a visiting professor at Defense Acquisition University. Title 17 U.S.C. 105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 U.S.C. 101 defines US Government work as work prepared by a military service member or employee of the US Government as part of that person's official duties.

Data Availability

Our deidentified data and maps can be found at [76]

Authors' Contributions

DB is the sole author and editor of this manuscript. He compiled and analyzed the data.

Conflicts of Interest

None declared.

References

1. Place RJ. Evaluation of the TRICARE program: fiscal year 2022 report to Congress. Health.mil. URL: https://www.health.mil/Reference-Center/Reports/2022/10/07/Evaluation-of-the-TRICARE-Program-FY-2022-Report-to-Congress [accessed 2024-03-14]
2. Audit of the Defense Health Agency's reporting of improper payment estimates for the Military Health Benefits Program. US Department of Defense. 2022 Jan 11. URL: https://media.defense.gov/2022/Jan/13/2002921501/-1/-1/1/DODIG-2022-052.PDF [accessed 2024-03-15]
3. Contract Resource Management Agency: financial report: fiscal year 2022. Health.mil. URL: https://health.mil/Reference-Center/Reports/2022/11/16/CRM-AFR-FY-22 [accessed 2024-03-14]
4. Medicare fraud & abuse: prevent, detect, report. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/Fraud-Abuse-MLN4649244.pdf [accessed 2024-03-14]
5. Berrada JA. A call for pre-exclusion notice and opportunity to respond. Public Contract Law J 2018;48(1):165-188 [FREE Full text]
6. Kumaraswamy N, Markey MK, Ekin T, Barner JC, Rascati K. Healthcare fraud data mining methods: a look back and look ahead. Perspect Health Inf Manag 2022 Jan 1;19(1):1i. [Medline: 35440932]
7. Levinson DR. Adverse events in hospitals: national incidence among Medicare beneficiaries. HW USA's Newsletter. URL: https://www.hwusa-newsletter.org/Newsletter-Pages/Images-2016/oei-06-09-00090.pdf [accessed 2023-12-01]
8. Burton B, Sun D, Jesilow P, Pontell HN. Two paths, one destination: a demographic portrait of physicians sanctioned by the federal government. J Health Hum Services Adm 2022;45(3):142-180 [FREE Full text]
9. Dunlap B, Basye A, Skillman SM. Background checks and the health workforce: practices, policies and equity. UW Department of Family Medicine. 2021 Nov. URL: https://familymedicine.uw.edu/chws/wp-content/uploads/sites/5/2021/11/Background-Checks-FR-2021.pdf [accessed 2023-12-05]
10. Medicaid Fraud Control Units. HHS Office of Inspector General. URL: https://oig.hhs.gov/fraud/medicaid-fraud-control-units-mfcu/ [accessed 2023-12-06]

11. GAO-09-174, excluded parties list system: suspended and debarred businesses and individuals improperly receive federal funds. US Government Accountability Office. 2009 Feb 25. URL: https://www.gao.gov/assets/a286497.html [accessed 2023-12-05]
12. Garlick C. Putting federal suspension & debarment officials in the driver's seat: empowering S&D programs to efficiently save taxpayer dollars. Public Contract Law J 2019;48(2):315-336 [FREE Full text]
13. Pachter J, Yukins C, Tillipman J. US debarment: an introduction. In: van Rooij B, Sokol DD, editors. The Cambridge Handbook of Compliance: Cambridge University Press; 2021:288 URL: https://scholarship.law.gwu.edu/faculty_publications/1711/
14. Interagency Suspension and Debarment Committee (ISDC). Acquistion.GOV. URL: https://www.acquisition.gov/isdc-compelling-reasons [accessed 2023-12-05]
15. 9.404 exclusions in the system for award management. Acquistion.GOV. URL: https://www.acquisition.gov/far/9.404 [accessed 2024-03-14]
16. Patel R, Sharma S. Credentialing: StatPearls Publishing; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK519504 [accessed 2023-12-01]
17. Health and Human Services Department. Updated special advisory bulletin on the effect of exclusion from participation in federal health care programs. Federal Register. 2013 May 9. URL: https://www.federalregister.gov/d/2013-11055 [accessed 2024-03-14]
18. Excluded Prescriber Monitoring Service: OIG audit risk mitigation for pharmacies. McKesson. URL: https://www.mckesson.com/Pharmacy-Management/Excluded-Prescriber-Monitoring/ [accessed 2023-12-01]
19. Find a claims address. TRICARE. URL: https://tricare.mil/GettingCare/PatientResources/contactus/claimsaddresses [accessed 2024-03-14]
20. Security Risk Assessment Tool. HealthIT.gov. URL: https://www.healthit.gov/topic/privacy-security-and-hipaa/security-risk-assessment-tool [accessed 2023-12-05]
21. 2021 HIMSS healthcare cybersecurity survey. Healthcare Information and Management Systems Society. 2021. URL: https://www.himss.org/sites/hde/files/media/file/2022/01/28/2021_himss_cybersecurity_survey.pdf [accessed 2024-03-14]
22. Ogu EC. Cybersecurity for eHealth: A Simplified Guide to Practical Cybersecurity for Non-Technical Healthcare Stakeholders & Practitioners: Routledge; 2021.
23. Personnel security clearances: additional actions needed to implement key reforms and improve timely processing of investigations. US Government Accountability Office. 2018 Mar 7. URL: https://www.gao.gov/products/gao-18-431t [accessed 2024-03-14]
24. Department of Defense. DoD patient bill of rights and responsibilities in the military health system (MHS). Washington Headquarters Services. 2011 Sep 26. URL: https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/600014p.pdf?ver=2020-04-03-132320-747 [accessed 2024-03-14]
25. Newman J, Ritchie A. Active duty servicemembers are more likely to report identity theft than other adults, new FTC data shows. Federal Trade Commission. 2020 May 21. URL: https://www.ftc.gov/news-events/news/press-releases/2020/05/active-duty-servicemembers-are-more-likely-report-identity-theft-other-adults-new-ftc-data-shows [accessed 2024-03-14]
26. Nicholas LH, Hanson C, Segal JB, Eisenberg MD. Association between treatment by fraud and abuse perpetrators and health outcomes among medicare beneficiaries. JAMA Intern Med 2020 Jan 1;180(1):62-69. [doi: 10.1001/jamainternmed.2019.4771] [Medline: 31657838]
27. Breach portal: notice to the secretary of HHS breach of unsecured protected health information. US Department of Health and Human Services: Office for Civil Rights. URL: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf [accessed 2023-12-01]
28. Fugitives. HHS Office of Inspector General. 2020 Dec 9. URL: https://oig.hhs.gov/fraud/fugitives/ [accessed 2023-12-01]
29. Consolidated Screening List. International Trade Administration. URL: https://www.trade.gov/consolidated-screening-list [accessed 2023-12-01]
30. Capitol violence. Federal Bureau of Investigation. 2021 Jan 8. URL: https://www.fbi.gov/wanted/capitol-violence [accessed 2023-12-01]
31. Federal & state exclusion sources. ProviderTrust. 2022 Apr 14. URL: https://www.providertrust.com/exclusion-sources/ [accessed 2023-12-01]
32. FDA debarment list (drug product applications). US Food and Drug Administration. 2023 Nov 1. URL: https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/compliance-actions-and-activities/fda-debarment-list-drug-product-applications [accessed 2023-12-01]
33. Excluded providers. Health.mil. URL: https://health.mil/About-MHS/OASDHA/Defense-Health-Agency/DHA-Office-of-the-Inspector-General/Fraud-and-Abuse/Excluded-Providers [accessed 2023-12-01]
34. Contract. US Department of Defense. 2022 Dec 22. URL: https://www.defense.gov/News/Contracts/Contract/Article/3254039/ [accessed 2023-12-05]
35. Data Analysis Tool. National Practitioner Data Bank. URL: https://www.npdb.hrsa.gov/analysistool/ [accessed 2023-12-05]
36. Find shortage areas. data.HRSA.gov. URL: https://data.hrsa.gov/tools/shortage-area [accessed 2023-12-01]

37. What is shortage designation? Bureau of Health Workforce. URL: https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation [accessed 2023-12-05]
38. Lewis VA, Spivack S, Murray GF, Rodriguez HP. FQHC designation and safety net patient revenue associated with primary care practice capabilities for access and quality. J Gen Intern Med 2021 Oct;36(10):2922-2928. [doi: 10.1007/s11606-021-06746-0] [Medline: 34346005]
39. Nakamura Y, Laberge M, Davis A, Formoso A. Barriers and strategies for specialty care access through federally qualified health centers: a scoping review. J Health Care Poor Underserved 2019;30(3):910-933. [doi: 10.1353/hpu.2019.0064] [Medline: 31422979]
40. Patient numbers by state. Health.mil. URL: https://www.health.mil/Military-Health-Topics/MHS-Toolkits/Media-Resources/Media-Center/Patient-Population-Statistics/Patient-Numbers-By-State [accessed 2023-12-05]
41. 2022 demographics profile data. Military OneSource. 2023 Nov 6. URL: https://www.militaryonesource.mil/data-research-and-statistics/military-community-demographics/2022-demographics-profile/ [accessed 2023-12-05]
42. Overview of the 2022 demographics profile dashboards. Military OneSource. URL: https://demographics.militaryonesource.mil/about [accessed 2024-02-23]
43. DoD personnel, workforce reports & publications. DMDC Web. URL: https://dwp.dmdc.osd.mil/dwp/app/dod-data-reports/workforce-reports [accessed 2024-02-22]
44. Shrestha A, Ho TE, Vie LL, et al. Comparison of cardiovascular health between US army and civilians. J Am Heart Assoc 2019 Jun 18;8(12):e009056. [doi: 10.1161/JAHA.118.009056] [Medline: 31164033]
45. Bond AM, Schwab SD. Utilization variation in military versus civilian care: evidence from TRICARE. Health Aff (Millwood) 2019 Aug;38(8):1327-1334. [doi: 10.1377/hlthaff.2019.00298] [Medline: 31381387]
46. Mulcahy AW, Phillips B, Whaley C. Balancing access and cost control in the TRICARE prescription drug benefit. RAND. URL: https://www.rand.org/content/dam/rand/pubs/research_reports/RR4400/RR4445/RAND_RR4445.pdf [accessed 2023-12-06]
47. Seshadri R, Strane D, Matone M, Ruedisueli K, Rubin DM. Families with TRICARE report lower health care quality and access compared to other insured and uninsured families. Health Aff (Millwood) 2019 Aug;38(8):1377-1385. [doi: 10.1377/hlthaff.2019.00274] [Medline: 31381389]
48. National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Board on Children, Youth, and Families, Committee on the Well-Being of Military Families. Demographic and military service characteristics of military families. In: Le Menestrel S, Kizer KW, editors. Strengthening the Military Family Readiness System for a Changing American Society: National Academies Press (US); 2019. URL: https://www.ncbi.nlm.nih.gov/books/NBK547615 [accessed 2023-12-05]
49. Shin P, Rosenbaum S, Richard Mauery D. Medicaid's role in treating children in military families. Himmelfarb Health Sciences Library: Health Sciences Research Commons. URL: https://hsrc.himmelfarb.gwu.edu/cgi/viewcontent.cgi?article=1171&context=sphhs_policy_facpubs [accessed 2023-12-05]
50. Provider directory. TRICARE West. URL: https://www.tricare-west.com/content/hnfs/home/tw/bene/provider-directory.html [accessed 2023-12-01]
51. Explore census data. US Census Bureau. URL: https://data.census.gov/ [accessed 2023-12-05]
52. Varma TN, Khan DA. MS Excel functions as supply chain fraud detector. Afr J Business Manage 2017 Feb 21;8(24) [FREE Full text]
53. Cheng C, Lee CC. A case study using data analytics to detect hail damage insurance claim fraud. J Forensic Accounting Res 2023 Dec 1;8(1):287-306 [FREE Full text] [doi: 10.2308/JFAR-2021-027]
54. DoD Instruction 6025.19: Individual Medical Readiness Program. Washington Headquarters Services. 2022 Jul 13. URL: https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/602519p.pdf?ver=R2OKfDXWHtMkQ9hrl8wdOw%3d%3d [accessed 2024-03-14]
55. Chen A, Blumenthal DM, Jena AB. Characteristics of physicians excluded from US Medicare and state public insurance programs for fraud, health crimes, or unlawful prescribing of controlled substances. JAMA Netw Open 2018 Dec 7;1(8):e185805. [doi: 10.1001/jamanetworkopen.2018.5805] [Medline: 30646294]
56. Zhang X, Lin D, Pforsich H, Lin VW. Physician workforce in the United States of America: forecasting nationwide shortages. Hum Resour Health 2020 Feb 6;18(1):8. [doi: 10.1186/s12960-020-0448-3] [Medline: 32029001]
57. Data Analysis Tool. National Practitioner Data Bank. URL: https://www.npdb.hrsa.gov/analysistool/ [accessed 2024-02-22]
58. Zecevic AA, Li AT, Ngo C, Halligan M, Kothari A. Improving safety culture in hospitals: facilitators and barriers to implementation of systemic falls investigative method (SFIM). Int J Qual Health Care 2017 Jun 1;29(3):371-377. [doi: 10.1093/intqhc/mzx034] [Medline: 28340250]
59. Public Use Data File. National Practitioner Data Bank. URL: https://www.npdb.hrsa.gov/resources/publicData.jsp [accessed 2023-12-01]
60. False Claims Act settlements and judgments exceed $2 billion in fiscal year 2022. US Department of Justice. 2023 Feb 7. URL: https://www.justice.gov/opa/pr/false-claims-act-settlements-and-judgments-exceed-2-billion-fiscal-year-2022 [accessed 2023-12-01]
61. Tucker D. The importance of a HIPAA security risk analysis. J Med Pract Manage 2022;38(3):103-104.

62. Schuster CR, Fossett C, Phillips G, Ellis J. DOD personnel: more accurate estimate of overdue security clearance reinvestigations is needed. Defense Technical Information Center. 2000 Sep 20. URL: https://apps.dtic.mil/sti/citations/trecms/AD1180286 [accessed 2023-12-06]
63. Kierkegaard P. Medical data breaches: notification delayed is notification denied. Computer Law Secur Rev 2012 Apr;28(2):163-183. [doi: 10.1016/j.clsr.2012.01.003]
64. Chandramouli R, Butcher Z. A zero trust architecture model for access control in cloud-native applications in multi-location environments. NIST Technical Series Publications. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207A.pdf [accessed 2024-03-15]
65. Zero Trust Maturity Model. Cybersecurity & Infrastructure Security Agency. URL: https://www.cisa.gov/resources-tools/resources/zero-trust-maturity-model [accessed 2023-12-01]
66. Frequently asked questions. Defense Finance and Accounting Service. URL: https://www.dfas.mil/militarymembers/faqs/ [accessed 2024-02-22]
67. TFA & online security. Defense Finance and Accounting Service. URL: https://www.dfas.mil/mypayinfo/2FA/TFA-Online-Security/ [accessed 2024-02-22]
68. The Privacy Act of 1974: 5 USC 552a as Amended: Reference Material: Internal Revenue Service; 1999.
69. Caralli RA, Allen JH, White DW. CERT Resilience Management Model (CERT-RMM): A Maturity Model for Managing Operational Resilience: Addison-Wesley Professional; 2010.
70. Malatji M. Management of enterprise cyber security: a review of ISO/IEC 27001:2022. In: 2023 International Conference On Cyber Management And Engineering (CyMaEn): IEEE:117-122. [doi: 10.1109/CyMaEn57228.2023.10051114]
71. Privacy framework. National Institute of Standards and Technology. 2018 Aug 10. URL: https://www.nist.gov/privacy-framework [accessed 2023-12-01]
72. Cybersecurity framework. National Institute of Standards and Technology. 2013 Nov 12. URL: https://www.nist.gov/cyberframework [accessed 2023-12-01]
73. TRICARE Manuals. URL: https://manuals.health.mil/ [accessed 2023-12-05]
74. FOIA library. Health.mil. URL: https://health.mil/Reference-Center/FOIA-Documents?refVector=0000001000000000&refSrc=134 [accessed 2024-03-14]
75. Defense Health Agency. Defense Health Agency reaffirms award of TRICARE West region contract to TriWest Healthcare Alliance. Health.mil. 2023 Apr 20. URL: https://www.health.mil/News/Articles/2023/04/20/Defense-Health-Agency-Reaffirms-Award-of-TRICARE-West-Region-Contract-to-TriWest-Healthcare-Alliance [accessed 2023-12-05]
76. TRICARE West provider roster exclusions. Kaggle. 2023 Nov 30. URL: https://www.kaggle.com/datasets/dbychkov/tricare-west-provider-roster-exclusions [accessed 2023-12-01]

Abbreviations

CRNA: certified registered nurse anesthetist
CSL: Consolidated Screening List
DFAS: Defense Finance and Accounting Services
DHA: Defense Health Agency
DO: doctor of osteopathic medicine
DODIG: Department of Defense's Office of Inspector General
FAR: Federal Acquisition Regulation
FBI: Federal Bureau of Investigation
FDA: Food and Drug Administration
FOIA: Freedom of Information Act
FQHC: Federally Qualified Health Center
FUG: list of fugitives
GSA: General Services Administration
HHS: Department of Health and Human Services
HIPAA: Health Insurance Portability and Accountability Act
HPSA: health provider shortage area
HRSA: Health Research Services Administration
ISDC: Interagency Suspension and Debarment Committee
JHIE: Joint Health Information Exchange
LCP: licensed clinical psychologist
LEIE: List of Excluded Individuals and Entities
MCSC: managed care support contractor
MD: doctor of medicine
MN: master of nursing

MTF: military treatment facility
NIST: National Institute of Standards and Technology
NPI: National Provider Identification
OIG: Office of Inspector General
PHI: personal health information
PI: Program Integrity
PII: personal identifying information
PT: physical therapist
RBT: registered behavior technician
SRA: security risk assessment
TFMR: Total Force Medical Readiness

Edited by G Eysenbach, T Leung; submitted 25.08.23; peer-reviewed by Anonymous, Anonymous; revised version received 01.12.23; accepted 05.02.24; published 09.04.24.

Please cite as:
Bychkov D
Insider Threats to the Military Health System: A Systematic Background Check of TRICARE West Providers
JMIRx Med 2024;5:e52198
URL: https://xmed.jmir.org/2024/1/e52198
doi:10.2196/52198

Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018

Ali Hazim Mustafa[1,*], MBChB, HD CM; Hanan Abdulghafoor Khaleel[2,*], MBChB, MSc CM, DrPH; Faris Lami[3,*], MBChB, PhD

[1]

[2]

[3]

[*]all authors contributed equally

Corresponding Author:
Hanan Abdulghafoor Khaleel, MBChB, MSc CM, DrPH

Related Articles:

Companion article: https://preprints.jmir.org/preprint/54611

Companion article: https://www.medrxiv.org/content/10.1101/2023.11.13.23298496v1

Companion article: https://med.jmirx.org/2024/1/e60433

Companion article: https://med.jmirx.org/2024/1/e60394

Companion article: https://med.jmirx.org/2024/1/e60194

Companion article: https://med.jmirx.org/2024/1/e64958

Abstract

Background: Brucellosis is both endemic and enzootic in Iraq, resulting in long-term morbidity for humans as well as economic loss. No previous study of the spatial and temporal patterns of brucellosis in Iraq was done to identify potential clustering of cases.

Objective: This study aims to detect the spatial and temporal distribution of human brucellosis in Iraq and identify any changes that occurred from 2007 to 2018.

Methods: A descriptive, cross-sectional study was conducted using secondary data from the Surveillance Section at the Communicable Diseases Control Center, Public Health Directorate, Ministry of Health in Iraq. The trends of cases by sex and age group from 2007 to 2018 were displayed. The seasonal distribution of the cases from 2007 to 2012 was graphed. We calculated the incidence of human brucellosis per district per year and used local Getis-Ord G_i* statistics to detect the spatial distribution of the data. The data were analyzed using Microsoft Excel and GeoDa software.

Results: A total of 51,508 human brucellosis cases were reported during the 12-year study period, with some missing data for age groups. Human brucellosis persisted annually in Iraq across the study period with no specific temporal clustering of cases. In contrast, spatial clustering was predominant in northern Iraq.

Conclusions: There were significant differences in the geographic distribution of brucellosis. The number of cases is the highest in the north and northeast regions of the country, which has borders with nearby countries. In addition, people in these areas depend more on locally made dairy products, which can be inadequately pasteurized. Despite the lack of significant temporal clustering of cases, the highest number of cases were reported during summer and spring. Considering these patterns when allocating resources to combat this disease, determining public health priorities, and planning prevention and control strategies is important.

(JMIRx Med 2024;5:e54611) doi:10.2196/54611

KEYWORDS

human brucellosis; livestock; clustering; spatial; temporal; Iraq

Introduction

Brucellosis is one of the most widespread zoonotic diseases in the world, responsible for enormous economic losses and considerable human morbidity in endemic areas [1]. The World Health Organization estimates that there are 500,000 annual new infections in over 170 countries, and 9% of them were from the Eastern Mediterranean region [2,3].

Most (90%) human infections are subclinical, resulting in delayed diagnosis and complications [1]. Human-to-human transmission is rare, and human infection can result from direct contact with infected animals or their products, the consumption of contaminated raw milk and milk products, or inhalation [1]. The main sources of brucellosis infection are most goats and sheep (*Brucella melitensis*), cattle (*Brucella abortus*), pigs (*Brucella suis*), and dogs (*Brucella canis*) [3]. Brucellosis seasonality shows no specific pattern. However, it usually coincides with the livestock breeding season [3] when human exposure to livestock or their contaminated products occurs, leading to more human infections in locations with infected livestock [4].

A study from northern Iraq showed that the prevalence of brucellosis in livestock varied from 1% to 70%, depending on the species and diagnostic methods [4]. That is, the seroprevalence of brucellosis in studies that used the Brewer card test was 3.1% for cattle, from 0.7% to 1% for sheep, from 2.5% to 4.4% for goats, and 2% for combined sheep and goat [5]. A study that used the Rose-Bengal test found a 5.5% prevalence in sheep and 5.3% in goats [6]. Studies that used competitive enzyme-linked immunosorbent assay (ELISA) found a brucellosis prevalence of 16.7% in cattle and 50% in buffalo [7]. The varying prevalence of brucellosis in livestock can have an effect on brucellosis prevalence in humans. In addition, the evaluation of veterinary and human public health measures can build on the finding of these studies. The veterinary vaccination program started in 2007 in Iraq; however, its implementation, like the implementation of all other health services and interventions, was negatively affected by political insecurities in the region, leading to suboptimal vaccination rates [4].

Currently, no satisfactory vaccine is available for humans [8]. Brucellosis control depends on testing and isolating or slaughtering brucellosis-positive animals, vaccinating susceptible animals, and controlling animal movements [9].

In Iraq, there is no prior description of the spatiotemporal epidemiology of human brucellosis. This study uses official data from the Ministry of Health (MoH) to identify potential changes in the spatial and temporal occurrence of human brucellosis cases in Iraq from 2007 to 2018.

Methods

Study Design

This was a descriptive, retrospective study of the spatial and temporal distribution of human brucellosis from 2007 to 2018.

Data Source

Human brucellosis data were extracted from the surveillance database at the Surveillance Section at the Communicable Diseases Control Center (CDC), Public Health Directorate, MoH in Iraq. The data included the number of human cases classified by the reporting districts for all 18 provinces in the country. The data for the study included information on sex, age group, the reporting district, and time of diagnosis (year). The number of human brucellosis cases in Iraq from 2007 to 2012 was retrieved from the Iraq CDC as the total number of patients at the provincial level; therefore, no further spatial analysis was done. From 2012 to 2018, the data were aggregated by province, district, age group, and sex; therefore, spatial analysis was done at the district level. Patients' sex was grouped as female and male. Age was grouped as younger than 1, 1 - 4, 5 - 14, 15 - 45, and older than 45 years in the aggregated form.

Up to 2018, all 18 Iraqi governorates must report human cases of brucellosis to the Surveillance Section using monthly aggregated forms by district. The presumptive diagnosis was made by the rapid brucellosis test, whereas the confirmatory test was done using ELISA or polymerase chain reaction test.

Data of each governorate's total population in Iraq and the population distribution by age and sex were retrieved from the Central Statistical Organization, Ministry of Planning, Iraq and used to calculate the incidence for the studied years (2007 - 2018). The incidence of human brucellosis was calculated as follows:

Incidence = number of new cases of human brucellosis in 1 year / district population for the same year × 100,000

No data about the population in each district of Kurdistan provinces for 2007, 2008, and 2009 were available from the Central Statistical Organization. Therefore, data from 2007 to 2012 were analyzed in graphs and tables using Excel software (2019; Microsoft). The 12-year study period was grouped into 2 parts: the first part spanned from 2007 to 2012 when the data were aggregated at the provincial level, and the second part spanned from 2013 to 2018 when the data were aggregated at the district level, to analyze and describe the spatial and temporal distribution of the disease.

Statistical Analysis

A descriptive analysis of human brucellosis during the first period was done by calculating the cases' frequency and percentage according to age group, sex, and season. In addition, we described the trends of sex and age group distribution from 2007 to 2018 using a stacked 100% bar chart and table to detect any changes during this period.

We used local Getis-Ord G_i* and Getis-Ord G_i statistics to identify the local concentration of high and low values in neighboring districts and their the statistical significance. The Getis-Ord G_i statistic is the ratio of the weighted average of the values in the neighboring locations to the sum of all values, not including the value at the location (x_i) [10]:

$$Gi=\sum j\neq iwijxj\sum j\neq ixj$$

In contrast, the local Getis-Ord G_i* statistic includes the value x_i in both the numerator and denominator:

$$Gi* = \sum_j w_{ij} x_j \Big/ \sum_j x_j$$

High values of either the local Getis-Ord G_i or Getis-Ord G_i* coefficient point to a concentration of districts with a high number of brucellosis cases, whereas low values point to a clustering of districts with a low number of cases. *P* values of .05 and lower were considered statistically significant. If the *P* value is statistically significant and the *z*-score is positive, then the spatial distribution of high and low values in the data set is more spatially clustered than would be expected if the underlying spatial processes were truly random. In contrast, if the *P* value is statistically significant and the *z*-score is negative, then the spatial distribution of high and low values in the data set is more spatially dispersed than would be expected if the underlying spatial processes were truly random. A dispersed spatial pattern often reflects some type of competitive process: a feature with a high value repels other features with high values and vice versa. *P* values of .05 and lower that were considered statistically significant were grouped at 3 thresholds (.05, .01, and .001) as the levels of significance increase with the lowest *P* value. Maps and tests were done using GeoDa software (version 1.12.1.161; September 2018; Center for Spatial Data Science at the University of Chicago).

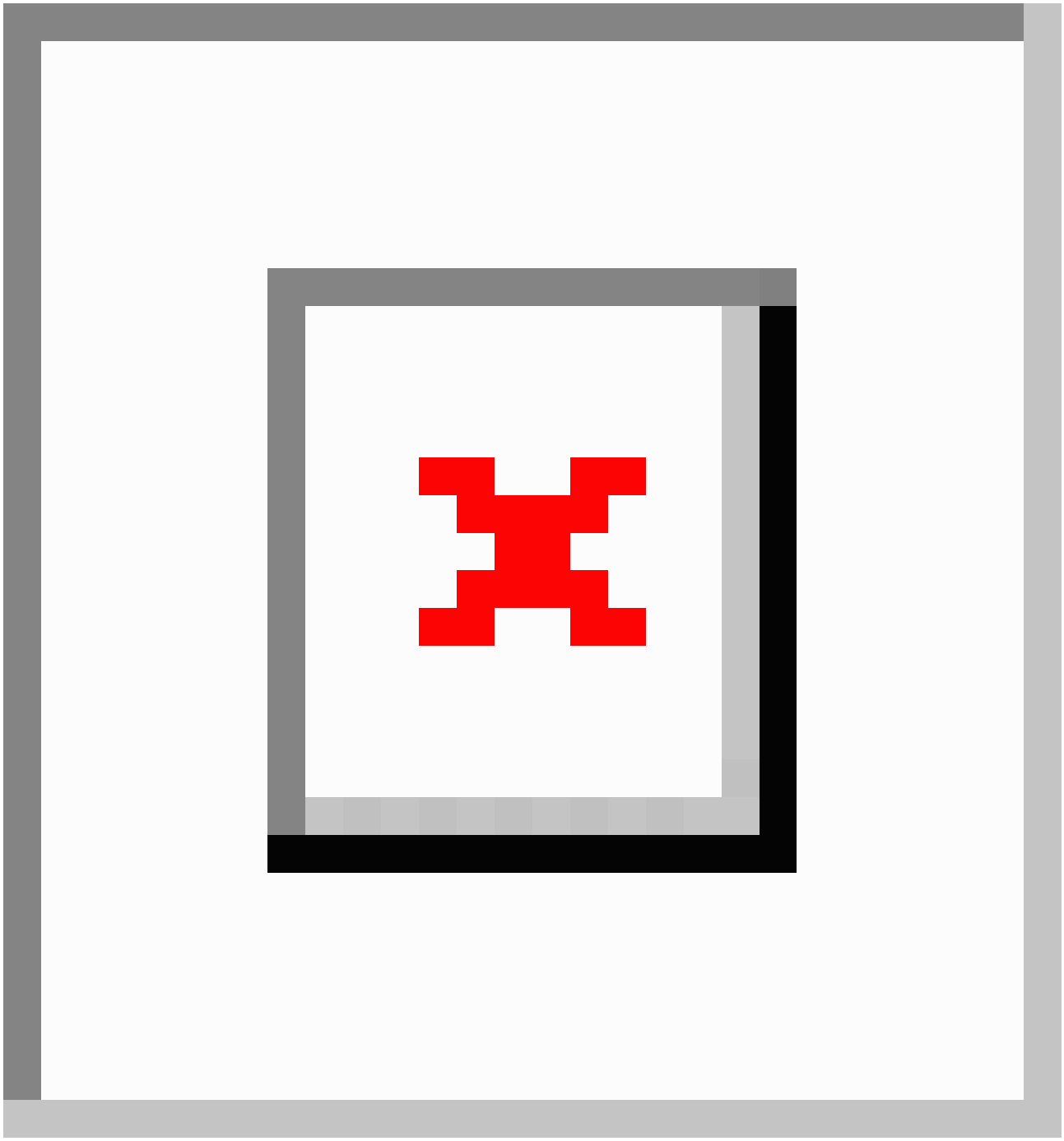
Ethical Considerations

Administrative and ethical approval was granted from the Public Health Directorate, MoH in Iraq (#5818).

Results

The total number of human brucellosis cases reported in Iraq from 2007 to 2018 was 51,508. The disease trend showed 2 peaks, one in 2010 and another in 2011, with 7399 and 7064 recorded cases, respectively. There was an apparent decline in reported cases from 2013 to 2018, as shown in Figure 1.

Figure 1. Frequency of reported human brucellosis cases in Iraq from 2007 to 2018.



Most cases were female (30,212/51,508, 58.02%), with increased frequency in 2016 and 2017, followed by decline in 2018 (Figure 2). Overall, 61.19% (30,977/50,621) of the cases were aged 15 - 45 years, with no apparent change in the age groups affected throughout the years (some data were missing for age groups; Table 1).

Figure 2. Sex distribution of human brucellosis cases in Iraq from 2007 to 2018.

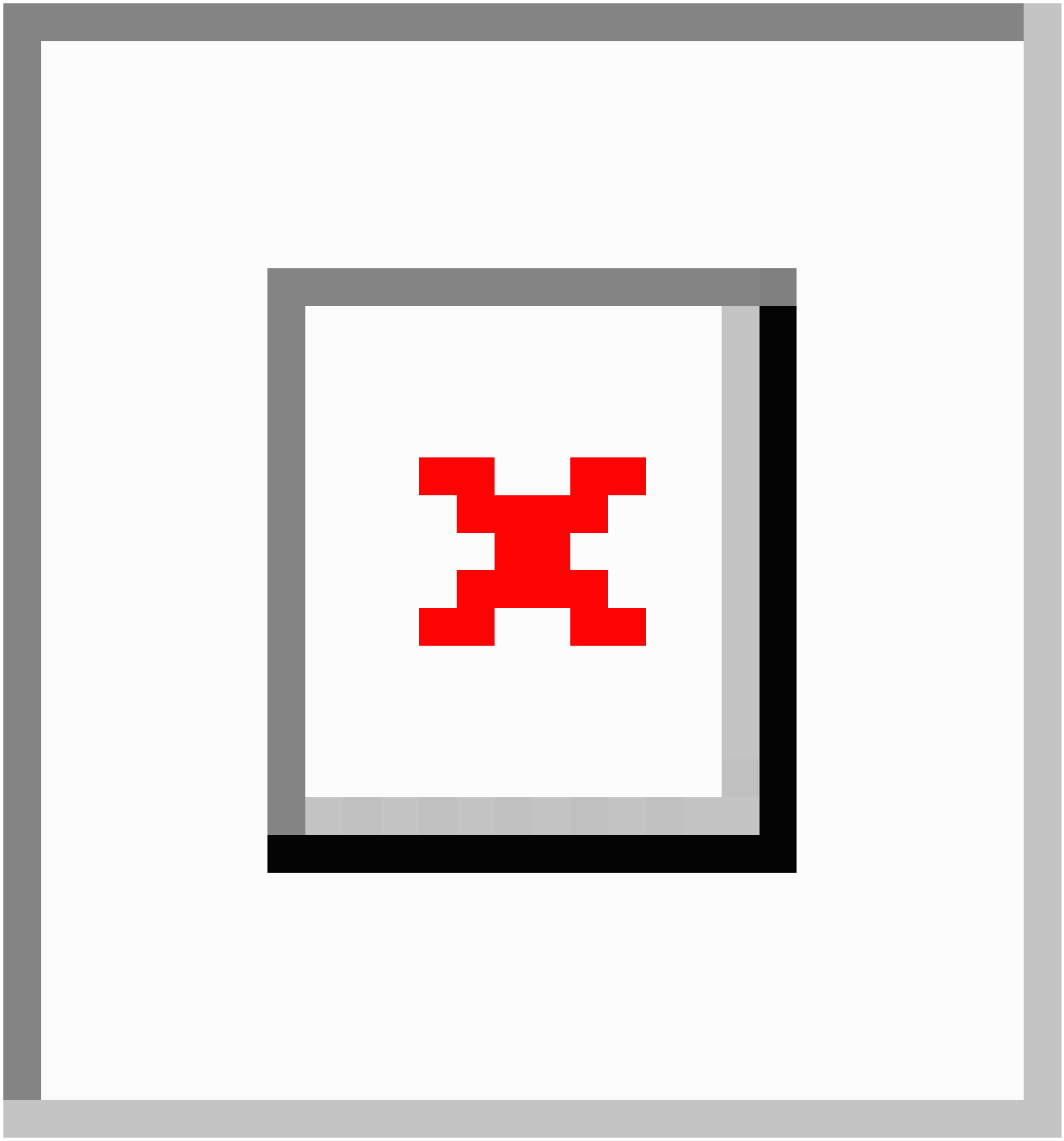
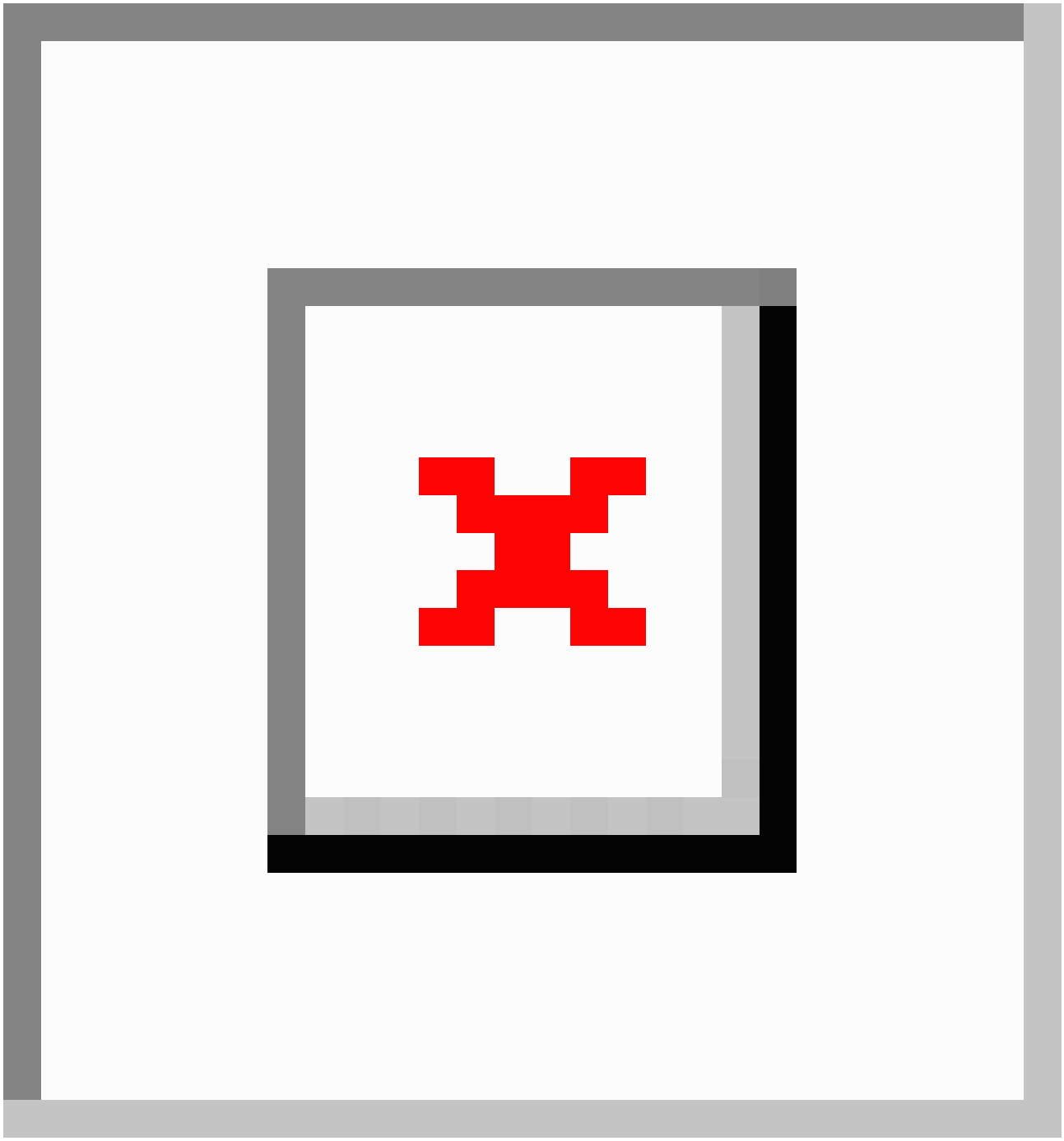


Table . Age group distribution of human brucellosis cases in Iraq from 2007 to 2018. Some data were missing for age groups.

Age group (y)	Year and cases, n (%)											
	2007 (n=6492)	2008 (n=6958)	2009 (n=6804)	2010 (n=7399)	2011 (n=7061)	2012 (n=5067)	2013 (n=3069)	2014 (n=2643)	2015 (n=2081)	2016 (n=1252)	2017 (n=1004)	2018 (n=791)
>45	1356 (20.89)	1788 (25.7)	1778 (26.13)	1608 (21.73)	1650 (23.37)	1080 (21.31)	679 (22.12)	607 (22.97)	421 (20.23)	252 (20.13)	210 (20.92)	165 (20.86)
15 - 45	3868 (59.58)	3905 (56.12)	3937 (57.86)	4744 (64.12)	4250 (60.19)	3092 (61.02)	1993 (64.94)	1698 (64.25)	1445 (69.44)	879 (70.21)	668 (66.53)	498 (62.96)
5 - 14	1099 (16.93)	1148 (16.5)	931 (13.68)	938 (12.68)	1000 (14.16)	766 (15.12)	358 (11.67)	280 (10.59)	193 (9.27)	111 (8.87)	110 (10.96)	113 (14.29)
1 - 4	159 (2.45)	108 (1.55)	150 (2.2)	92 (1.24)	148 (2.1)	112 (2.21)	36 (1.17)	55 (2.08)	20 (0.96)	8 (0.64)	14 (1.39)	14 (1.77)
<1	10 (0.15)	9 (0.13)	8 (0.12)	17 (0.23)	13 (0.18)	17 (0.34)	3 (0.1)	3 (0.11)	2 (0.1)	2 (0.16)	2 (0.2)	1 (0.13)

Seasonal distribution of human brucellosis cases was constant throughout the year (Figure 3). The cases occurred the most frequently in summer (15,044/39,794, 37.8%), followed by spring (10,938/39,794, 27.49%).

Figure 3. Seasonal distribution of human brucellosis cases in Iraq from 2007 to 2012.



The local Getis-Ord G_i* statistics showed increased hot spots (high-high clusters) and cold spots (low-low clusters) from 2013 to 2018. Hot spots were located in the north and northeastern parts of Iraq, that is, in districts Choman, Soran, and Erbil in Erbil province; district Raniya in Sulaymaniyah province; district Akre in Nineveh province; districts Kirkuk and Dibis in Kirkuk province; districts Alshirqat, Baiji, and Tikrit in Salah Al-Din province; and district Amedi in Dahuk province. In contrast, cold spots were located in districts located in the provinces of Thiqar, Muthanna, Maysan, Anbar, Najaf, and Baghdad. District Dibis in Kirkuk shifted from a cold spot in 2015 to a hot spot in 2016. District Koisanjaq in Erbil shifted from a hot spot in 2014 to a cold spot in 2015 (Figures 4-10).

Figure 4. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2013.

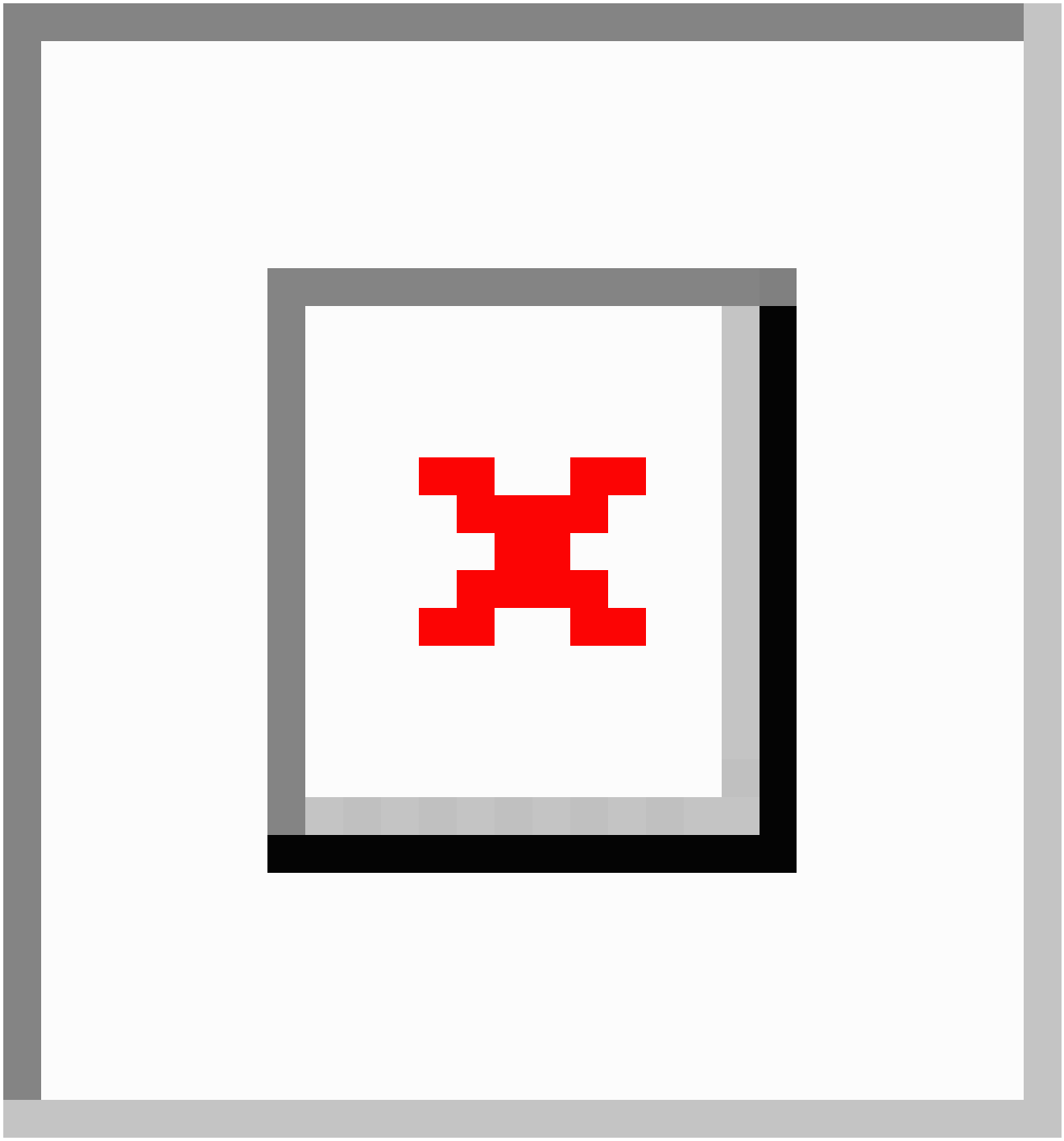


Figure 5. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2014.

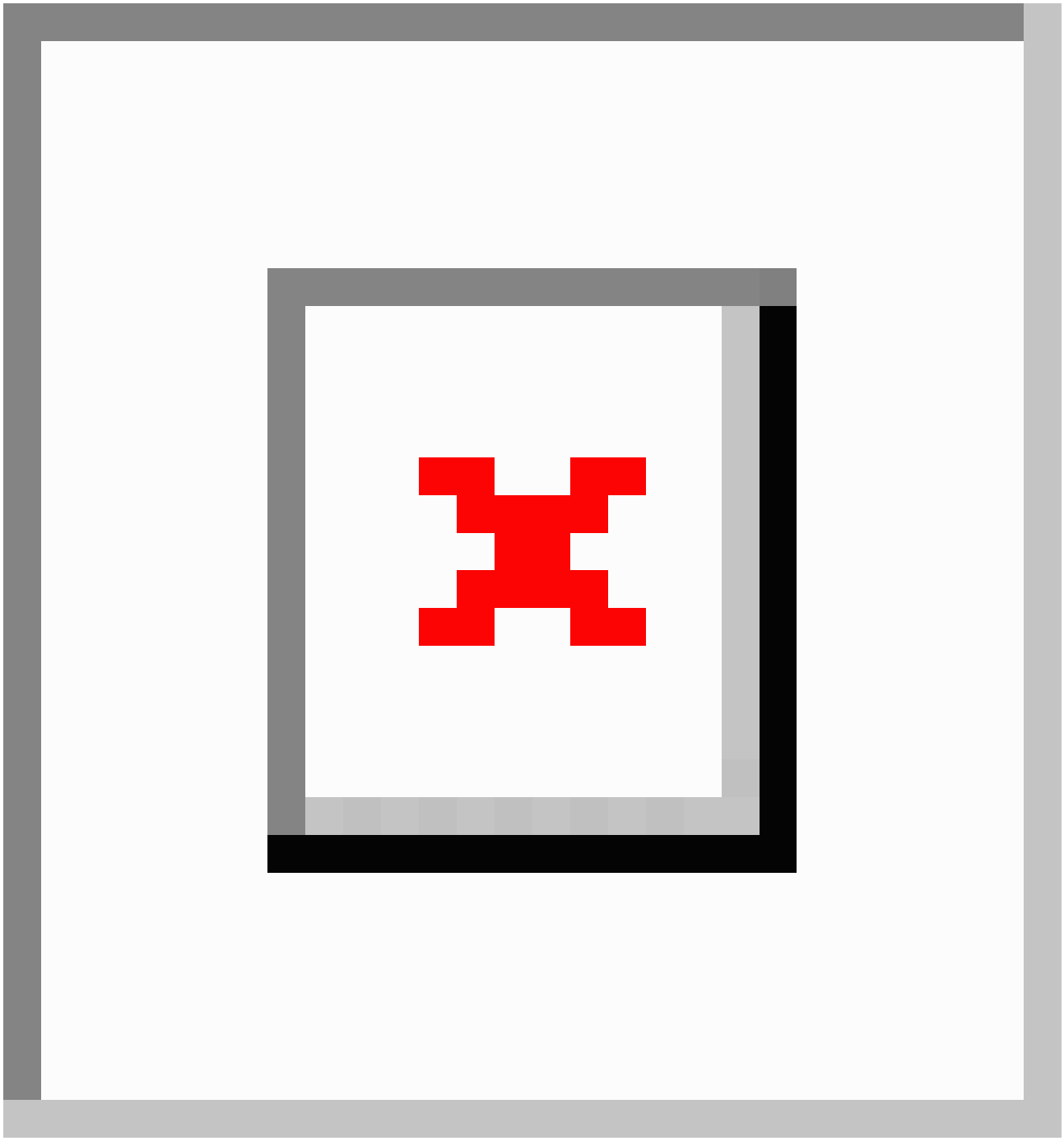


Figure 6. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2015.

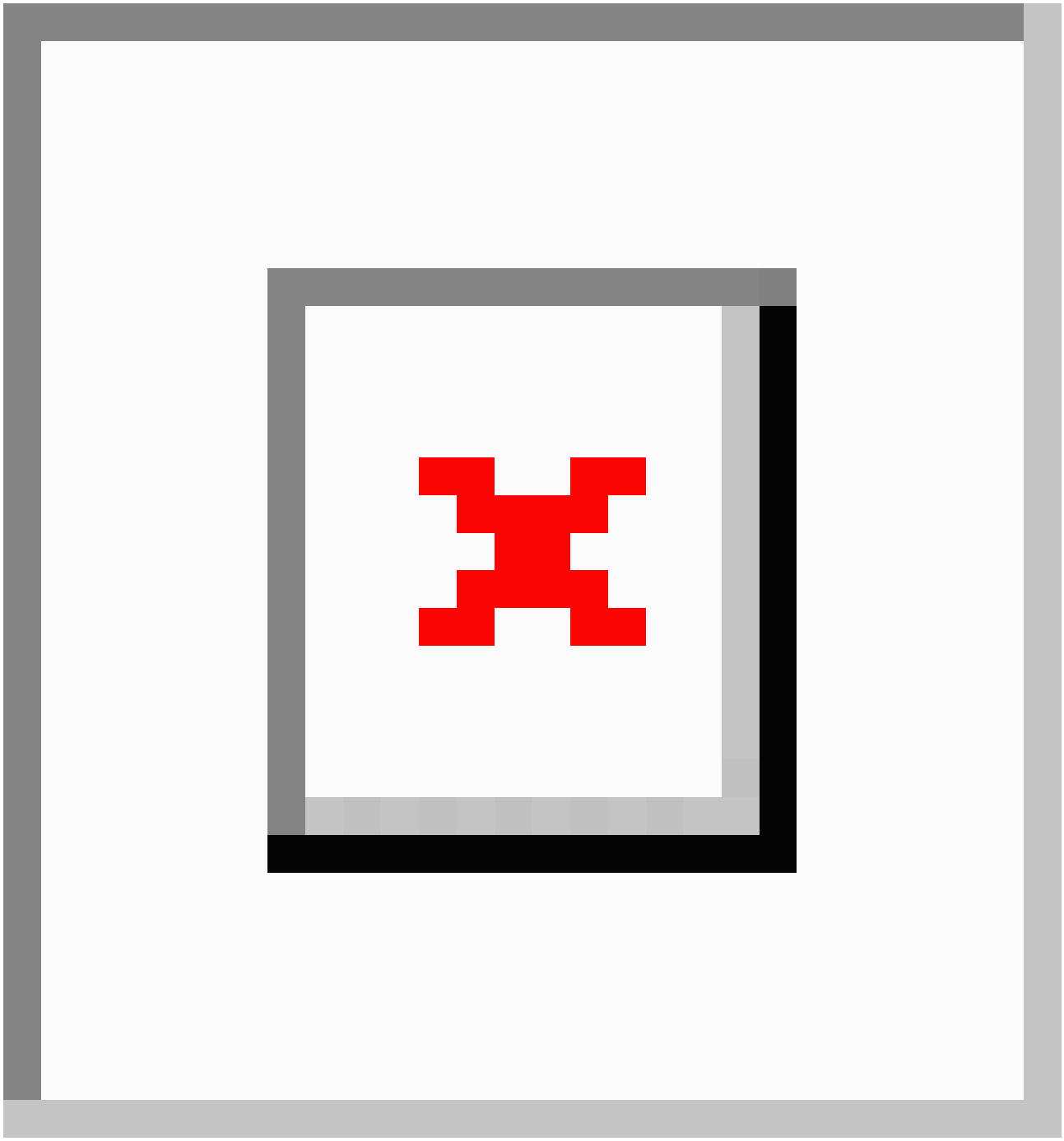


Figure 7. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2016.

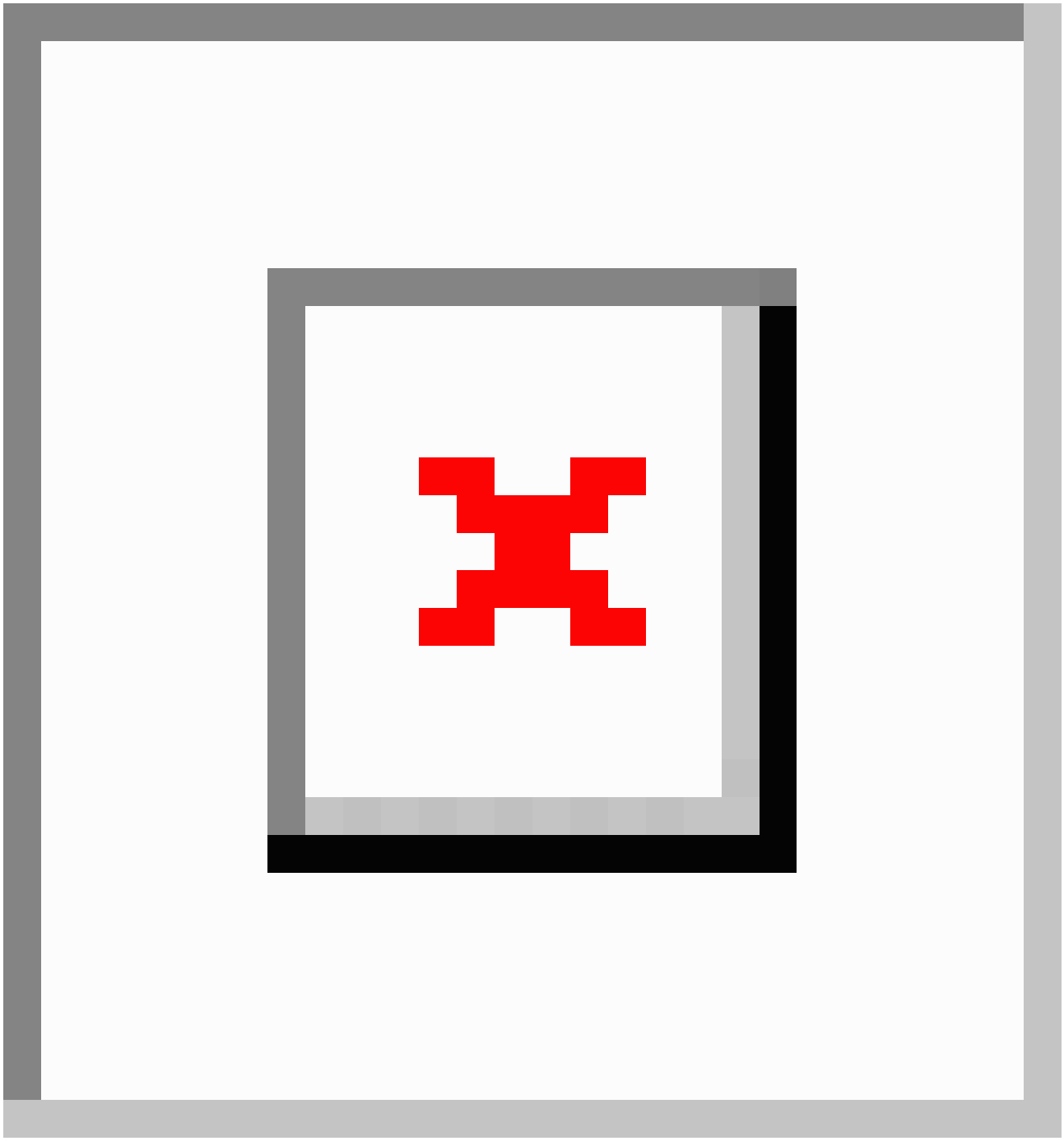


Figure 8. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2017.

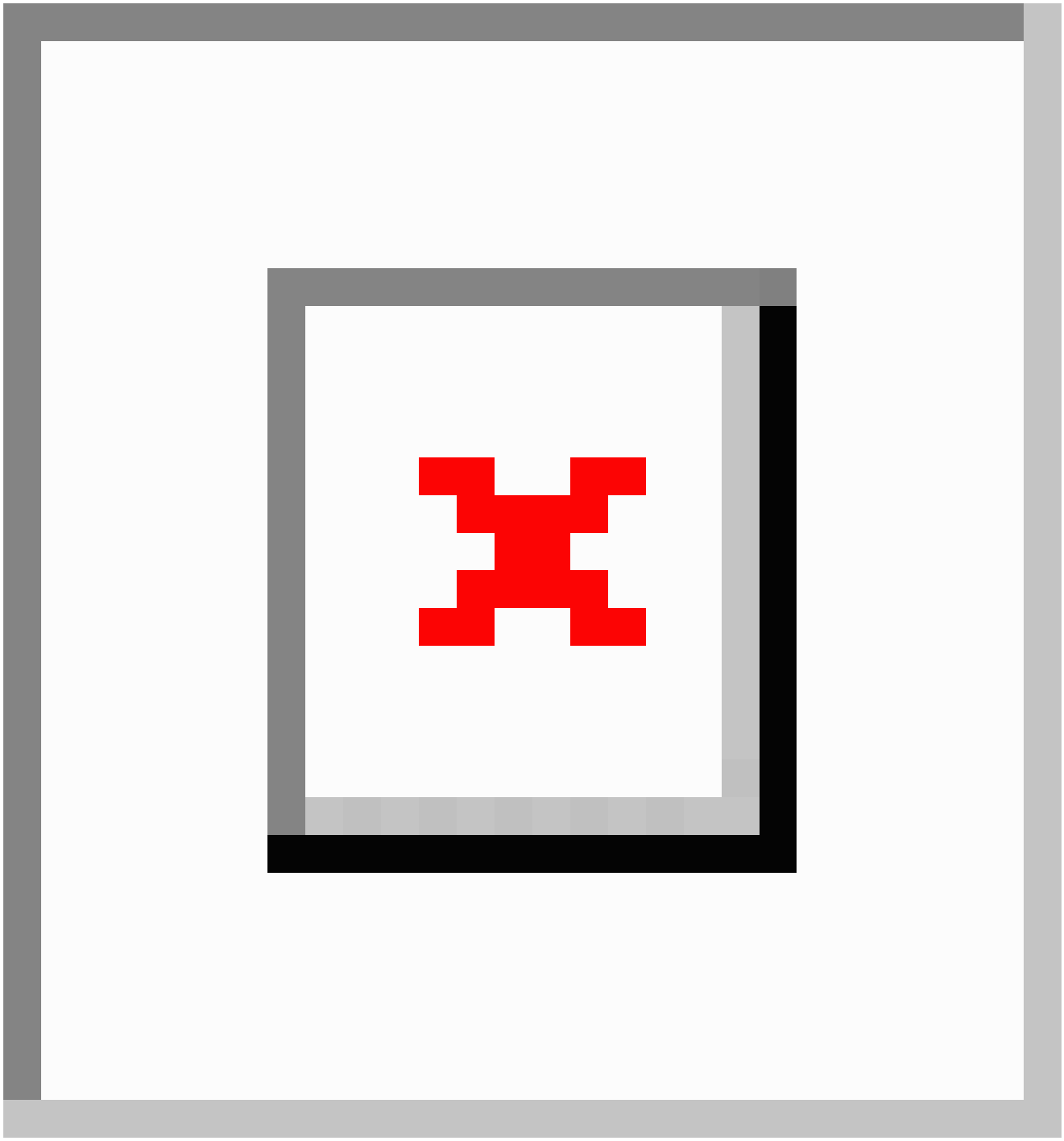


Figure 9. Gi* Cluster (A) and Gi* Significance (B) map of human brucellosis cases in Iraq in 2018.

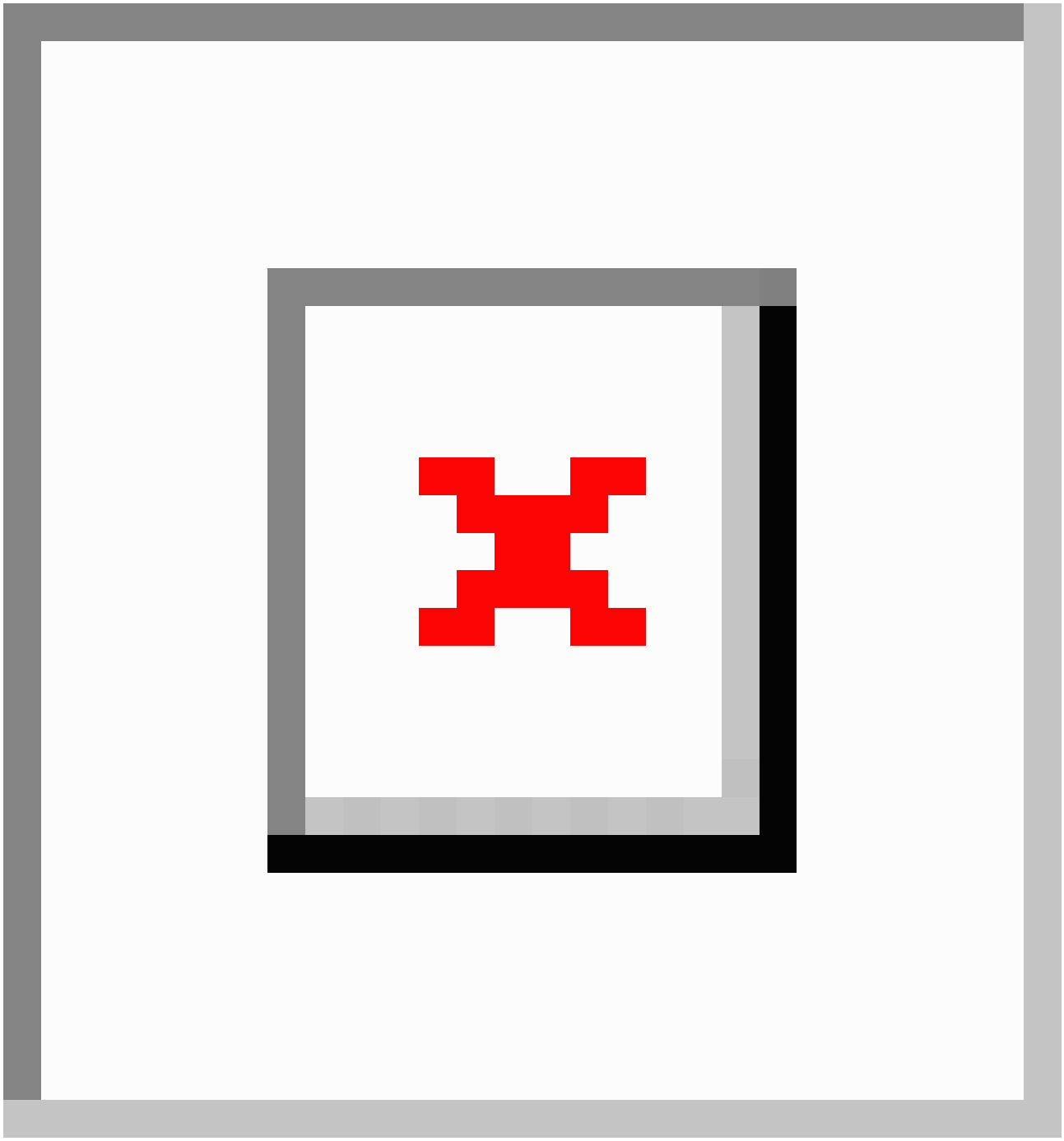
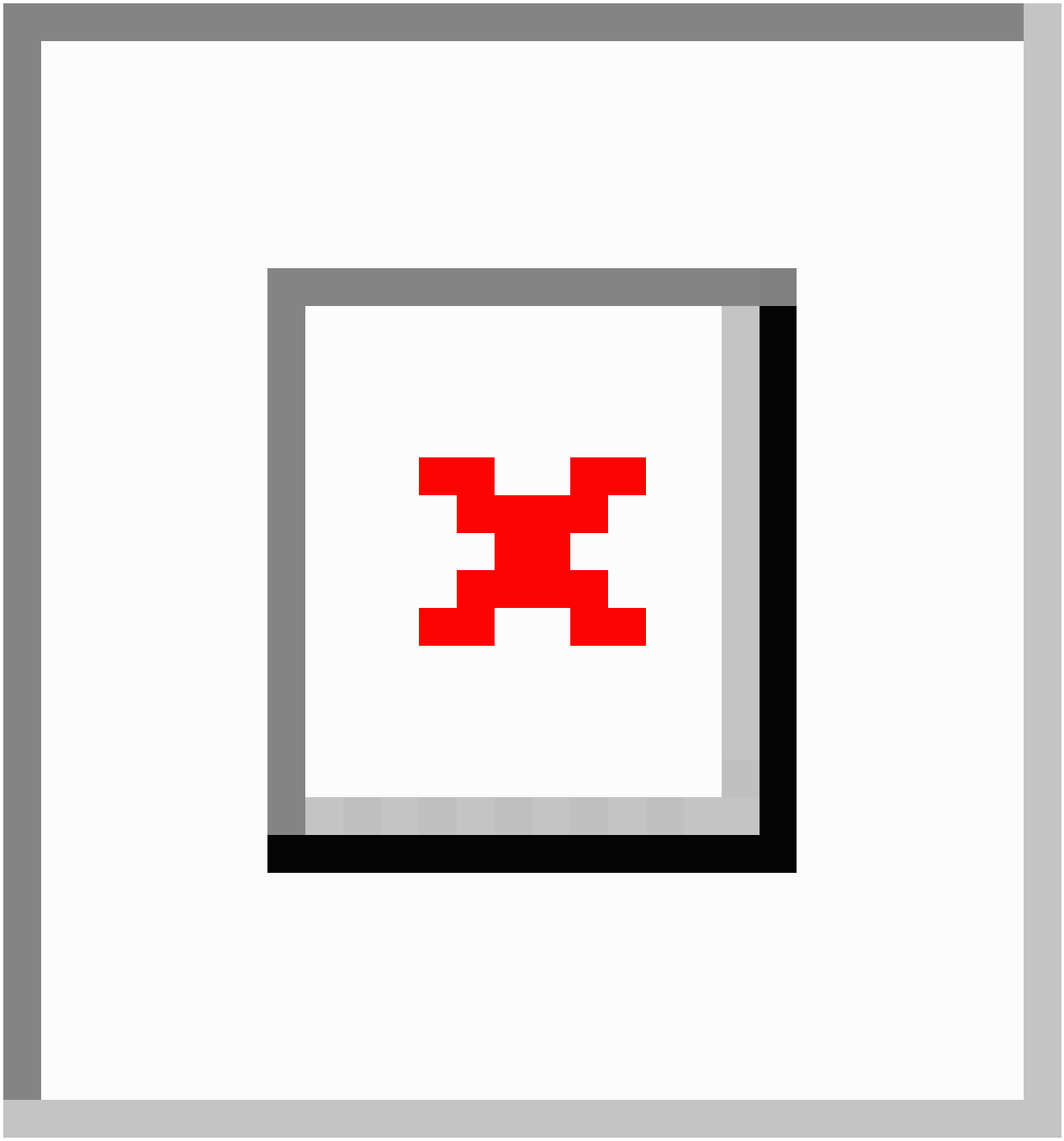


Figure 10. Spatial distribution maps of human brucellosis incidence rate in 2013 (A) and 2018 (B).



Discussion

Principal Findings

Analyzing the spatiotemporal distribution of human brucellosis is important in understanding the epidemiology of the disease in the country and helps forecast the future situation. There has been a decline in human brucellosis cases since 2012, which may be due to veterinary protective and control measures through vaccinating livestock that have been implemented since 2007. However, vaccination rounds were disrupted during Islamic State in Iraq and the Levant (ISIL) events that negatively affected the control program [11,12]. Nevertheless, animal health campaigns were relaunched soon after liberation in 2017, with several vaccination rounds having been completed since then [13].

The number of female individuals who tested positive for brucellosis has been constantly higher than the number of male individuals, with increased reporting frequency since 2015. This can be due to increased exposure of female individuals to infected animals and animal products through housekeeping and farming activities [11]. This contradicts what has been found in research from Germany [1], Italy [14], and Iran [8]. Therefore, the potential differences should be investigated further to detect whether the difference is due to different proportions of male individuals in different countries or biological differences that enhance contracting brucellosis. Nearly two-thirds (30,977/50,621, 61.19%) of the patients were between 15 and

45 years old, which can negatively affect the economic status of the patients and their families and the quality of life for the patients [9]. People in this age group may consume more locally made dairy products due to a lack of knowledge regarding the importance of pasteurized milk and milk products. However, this age category was very broad and could have been classified into 2-3 age groups to detect the most commonly affected age group [15-18].

Most human brucellosis cases were in the northern provinces during the first period (2007 - 2012), especially in Salah Al-Din, Sulaymaniyah, Nineveh, Erbil, Kirkuk, and Dahuk. Northern provinces share an extensive border with Iran, Turkey, and Syria and direct borders with other Iraq provinces, which, in turn, share borders with Jordan, Saudi Arabia, and Kuwait [16,17]. Sheep and goat husbandry has been practiced in the northern region of Iraq since the earliest times due to the nature of these provinces, which is characterized by broad grass-covered terrain, undulating hills steep, and craggy mountains well suited to sheep and goat grazing [19]. Livestock and its products dominate northern Iraq's economy and represents an important food resource for its inhabitants [5]. Therefore, the high clustering of brucellosis in this area could be due to area grazing strategy, animal population density, owner's ignorance of the hazards of the disease, unhygienic disposal of infected animals as well as aborted fetuses or placental membranes, and uncontrolled movement of diseased animals [18]. Because there is no curative medical therapy for animal brucellosis, the infected animals should be slaughtered and disposed of properly. However, these animals are expensive in Iraq; therefore, farmers firmly refuse to slaughter and dispose them. As a result, the infection will persist through transmission from the infected animals to their offspring by breastfeeding. Likewise, humans may consume this infected, unpasteurized milk, resulting in infection and areas endemic with brucellosis in animals and humans [15]. One potential control method in this case is the isolation of the infected animal from healthy animals and milk pasteurization to prevent animal-to-animal and animal-to-human transmissions. Preventive measures such as health education activities should be performed in high-risk areas. Adopting the Quarantine-Slaughter-Immunization strategy and One Health Approach is crucial in controlling the disease. This can be achieved through multisectoral coordination and coordination with neighboring countries in the control programs.

The seasonal variation in the occurrence of human brucellosis in Iraq, that is, the decline in the number of human brucellosis cases in winter and the rise in spring and summer, could be explained by high exposure to the disease during spring and summer when the deliveries of animals, increased milk production, and contamination occur [19,20].

The human brucellosis cases were clustered in the north and northeastern regions of Iraq, mainly in areas in the vicinity of the Zagros mountains, which contain dense oak forests and fertile soil with a high density of sheep and goats; the main economic activity in the area is animal farming. The Zagros Mountains are also the route of seasonal migration for nomads [19]. Human-to-human transmission is rare and clusters of human cases are most likely a result of animal processing, more intensive agricultural production zones, and similar sociocultural practices [14].

On the contrary, the low cluster areas may be due to having economic superiority, standardized industrial manufacturing for cow- or sheep-related products, good sanitary habits, awareness of human brucellosis, and easy access to immediate treatment after infection [4,21]. In addition, several changes in the surveillance case definition, the diagnostic methods, and the reporting of human brucellosis that occurred during the second period of the study could have also contributed to the low numbers reported compared to the first period. For example, the primary diagnostic method during the first period was the Rose-Bengal test, which has variable sensitivity and specificity based on the exposure history, stage of infection, and prior infection history. In contrast, the diagnostic test used during the second period was ELISA for immunoglobulin M, immunoglobulin G, and culture. Although all suspected cases are initially reported to the Surveillance Section, their final classification as suspected or confirmed may vary by governorates based on their testing and interpretation capacity (Surveillance Section, CDC, Public Health Directorate, MoH in Iraq; unpublished data; 2024). As another example, the data were collected in an aggregate format based on sex, age group, and province in the first period of the study. In contrast, the data were collected in a case-based format during the second period of the study (Surveillance Section, CDC, Public Health Directorate, MoH in Iraq; unpublished data; 2024).

This study has several strengths. First, this is the first study to include data across 12 years in a detailed spatiotemporal analysis. The consistency of the epidemiological characteristics across 12 years can lead to public health interventions in areas where effort should be spent. That is, due to the fact that the infection is consistently more frequent among female individuals, health promotion activities may be directed to design educational materials on dealing with animals and their products and delivering them using personal protective measures that target female individuals working in the animal husbandry industry. The second strength is the use of an innovative statistical approach to detect whether the clustering of cases is significant or random, in addition to follow-up of the progress of clusters over 7 years, which clearly showed a continued increase in the north and northeastern areas and a continuous decline in the southern areas.

This study also has limitations. Data regarding important risk factors of human brucellosis, such as occupation, rural or urban areas, comorbidities, and treatment protocol, were lacking. Had this information been available, it would have facilitated a better understanding of the epidemiology and characteristics of human brucellosis in Iraq (Surveillance Section, CDC, Public Health Directorate, MoH in Iraq; unpublished data; 2024).

Conclusions

Despite the declining incidence of human brucellosis in Iraq from 2013 to 2018, human brucellosis is still endemic and constitutes a public health problem. Most cases were reported in the summer and spring seasons among female individuals and those aged 15 - 45 years. Human brucellosis cases presented significant spatial clustering in northern and northeastern areas.

Conflicts of Interest

None declared.

References

1. Enkelmann J, Stark K, Faber M. Epidemiological trends of notified human brucellosis in Germany, 2006-2018. Int J Infect Dis 2020 Apr;93:353-358. [doi: 10.1016/j.ijid.2020.02.019] [Medline: 32081773]
2. Bagheri H, Tapak L, Karami M, Amiri B, Cheraghi Z. Epidemiological features of human brucellosis in Iran (2011-2018) and prediction of brucellosis with data-mining models. J Res Health Sci 2019 Dec 4;19(4):e00462. [Medline: 32291361]
3. Yang H, Zhang S, Wang T, et al. Epidemiological characteristics and spatiotemporal trend analysis of human brucellosis in China, 1950-2018. Int J Environ Res Public Health 2020 Mar 31;17(7):2382. [doi: 10.3390/ijerph17072382] [Medline: 32244493]
4. Dahl MO. Brucellosis in food-producing animals in Mosul, Iraq: a systematic review and meta-analysis. PLoS One 2020 Jul 9;15(7):e0235862. [doi: 10.1371/journal.pone.0235862] [Medline: 32645099]
5. Karim MA, Penjouian EK, Dessouky FI. The prevalence of brucellosis among sheep and goats in northern Iraq. Trop Anim Health Prod 1979 Aug;11(3):186-188. [doi: 10.1007/BF02237799] [Medline: 505594]
6. Al-Iraqi OM, Al-Hankawe OK, Abdul-Majeed MO, Al-Farwachi MI. Comparison between competitive ELISA and Rose-Bengal tests in detection of Brucella antibodies in buffalo sera in Mosul city, Iraq. Bas J Vet Res 2009 Jun;8(1):93-97 [FREE Full text]
7. Al-Hankawe OK, Al-Iraqi OM, Al-Farwachi MI, Abdul-Majeed MO. Using of competitive ELISA in detection of Brucella antibodies in cattle sera. Iraqi J Veterinary Sci 2009 Dec;23(2):97-103. [doi: 10.33899/ijvs.2009.5739]
8. Zeinali M, Doosti S, Amiri B, Gouya MM, Godwin GN. Trends in the epidemiology of brucellosis cases in Iran during the last decade. Iran J Public Health 2022 Dec;51(12):2791-2798. [doi: 10.18502/ijph.v51i12.11470] [Medline: 36742247]
9. Yu X, Fang M, Li Y, et al. Epidemiological characteristics and spatio-temporal analysis of brucellosis in Shandong province, 2015-2021. BMC Infect Dis 2023 Oct 9;23(1):669. [doi: 10.1186/s12879-023-08503-6] [Medline: 37814221]
10. Ord JK, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 1995 Oct;27(4):286-306. [doi: 10.1111/j.1538-4632.1995.tb00912.x]
11. Khalid HM. Seroprevalence and associated risk factors of brucellosis among human population in Duhok city, Iraq. Infect Drug Resist 2023 May 8;16:2805-2811. [doi: 10.2147/IDR.S407263] [Medline: 37187481]
12. Bagheri H, Tapak L, Karami M, et al. Forecasting the monthly incidence rate of brucellosis in west of Iran using time series and data mining from 2010 to 2019. PLoS One 2020 May 12;15(5):e0232910. [doi: 10.1371/journal.pone.0232910] [Medline: 32396582]
13. FAO. Iraq: FAO launches emergency animal health campaign in former ISIL-held Mosul area. ReliefWeb. 2017 Oct 23. URL: https://reliefweb.int/report/iraq/iraq-fao-launches-emergency-animal-health-campaign-former-isil-held-mosul-area [accessed 2023-11-05]
14. Facciolà A, Palamara MAR, D'Andrea G, et al. Brucellosis is a public health problem in Southern Italy: burden and epidemiological trend of human and animal disease. J Infect Public Health 2018;11(6):861-866. [doi: 10.1016/j.jiph.2018.07.007] [Medline: 30056067]
15. Al-Koubaisy HN, Lafi SA. Presentation of brucellosis in an endemic area; west of Iraq. Egypt Acad J Biol Sci 2011 Dec;3(1):13-18. [doi: 10.21608/eajbsg.2011.16690]
16. Abbas BA, Aldeewan AB. Occurrence and epidemiology of Brucella spp. in raw milk samples at Basrah province, Iraq. Bulgarian J Veterinary Med 2009;1(12):136-142 [FREE Full text]
17. Shareef JM. A review of serological investigations of brucellosis among farm animals and humans in northern provinces of Iraq (1974–2004). J Vet Med B 2006 Dec;53(s1):38-40. [doi: 10.1111/j.1439-0450.2006.01021.x]
18. Al-Samarrae EA. Detection of Brucella antibodies of sheep in Al-Anbar province by using some serological tests. Iraqi J Vet Med 2013 Jun 30;37(1):7-12. [doi: 10.30539/iraqijvm.v37i1.316]
19. Al-Khafaji JKT. Brucellosis among human populations in Al-Musaib district, Babylon province/Iraq. Al-Mustansiryia Sci J 2003;14(1):1-5 [FREE Full text]
20. Moosazadeh M, Abedi G, Kheradmand M, Safiri S, Nikaeen R. Seasonal pattern of brucellosis in Iran: a systematic review and meta-analysis. Iran J Health Sci 2016;4(1):62-72. [doi: 10.18869/acadpub.jhs.4.1.62]
21. Chen Q, Lai S, Yin W, et al. Epidemic characteristics, high-risk townships and space-time clusters of human brucellosis in Shanxi province of China, 2005-2014. BMC Infect Dis 2016 Dec 19;16(1):760. [doi: 10.1186/s12879-016-2086-x] [Medline: 27993134]

Abbreviations

CDC: Communicable Diseases Control Center
ELISA: enzyme-linked immunosorbent assay

ISIL: Islamic State in Iraq and the Levant
MoH: Ministry of Health

Edited by E Meinert; submitted 16.11.23; peer-reviewed by Anonymous, Anonymous; revised version received 07.04.24; accepted 29.04.24; published 03.07.24.

Please cite as:
Mustafa AH, Khaleel HA, Lami F
Human Brucellosis in Iraq: Spatiotemporal Data Analysis From 2007-2018
JMIRx Med 2024;5:e54611
URL: https://xmed.jmir.org/2024/1/e54611
doi:10.2196/54611

Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial

Sunday Udo[1], MPH, PhD; Pius Ogbu Sunday[1], MSc; Paul Alumbugu Tsaku[1], BSc, MSc, PhD; Israel Olaoluwa Oladejo[1], MPH; Anthony Meka[2], MBBS; Linda Chinonso Ugwu[2], MPH; Motunrayo Ajisola[3], PhD; Joshua Akinyemi[3], PhD; Abiola Oladejo[3], MSc; Akinyinka Omigbodun[3], PhD; Sopna Mannan Choudhury[4], MPH; Jo Sartori[4], MSc; Onaedo Ilozumba[4], PhD; Sam Watson[4], PhD; Richard Lilford[4], PhD

1

2

3

4

Corresponding Author:
Paul Alumbugu Tsaku, BSc, MSc, PhD

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.14.23292603v1

Companion article: https://med.jmirx.org/2024/1/e57310

Companion article: https://med.jmirx.org/2024/1/e56498

Companion article: https://med.jmirx.org/2024/1/e56442

Abstract

Background: Leprosy leads to nerve damage and slow-healing ulcers, which are treatable with routine therapy. There has been a recent resurgence of interest in the use of honey for the treatment of different kinds of wounds.

Objective: The aim of this study, Honey Experiment on Leprosy Ulcer (HELP), is to evaluate the healing properties of raw, unadulterated African honey in comparison with normal saline dressing for the treatment leprosy ulcers.

Methods: This is a multicenter, comparative, prospective, single-blinded, parallel-group, and 1:1 individually randomized controlled trial to be conducted at The Leprosy Referral Hospital, Chanchaga in Minna, Niger State, North Central Nigeria, and St. Benedict Tuberculosis and Leprosy Rehabilitation Hospital in Ogoja, Cross River State, South-South Nigeria. Raw, unadulterated honey will be used in the ulcer dressing of eligible, consenting participants in the intervention group, whereas those in the control group will be treated by dressing with normal saline. The main outcomes will be the proportion of complete healing and the rate of healing up to 84 days after randomization. Follow-up will be conducted 6 months after randomization. We aim to enroll 90 - 130 participants into the study. Blinded observers will examine photographs of ulcers to determine the outcomes.

Results: The recruitment of trial participants began on March 14, 2022, and has been continuing for approximately 24 months.

Conclusions: Our study will provide an unbiased estimate of the effect of honey on the healing of neuropathic ulcers.

Trial Registration: ISRCTN registry ISRCTN10093277; https://www.isrctn.com/ISRCTN10093277

(JMIRx Med 2024;5:e50970) doi:10.2196/50970

KEYWORDS

leprosy; ulcers; wounds; honey; neuropathy; nerves; Africa; randomized controlled trial; RCT

Introduction

Background and Rationale

Leprosy or Hansen disease is a chronic granulomatous disease caused by *Mycobacterium leprae* [1]. The disease morbidity is characterized by damage to the skin and peripheral nerves, causing neuropathy and severe disability and consequently resulting in social exclusion and stigmatization [2]. Leprosy is preventable and treatable with multidrug therapy but remains endemic in many communities of people living in poverty and with poor hygiene [2]. Nigeria reported a total of 1417 new cases of leprosy in 2020, including 87 new child cases [3], with a grade-2 disability rate of about 15% for the past 5 consecutive years [4].

A total of 30% to 50% of people infected with leprosy globally are reported to eventually have nerve damage or neuropathy [5]. Ulcers usually occur in anesthetic feet and will heal slowly with routine therapy; however, they have a tendency to recur [6].

The use of honey as a therapeutic agent in the treatment of wounds dates back to ancient times, with the earliest documented report being recorded in the *Edwin Smith Papyrus* (2600 - 2200 BCE) [7]. Honey is a viscous, supersaturated solution containing sugars, water, amino acids, vitamins, minerals, enzymes, and many other substances that is derived from nectar gathered and modified by the honeybee, *Apis mellifera* [8,9]. Studies have suggested that honey promotes wound healing, stimulates tissue growth, facilitates debridement and epithelization, deodorizes, reduces edema and exudates, and possesses antimicrobial properties [7,10,11].

There has been a recent resurgence of interest in the use of honey for the treatment of different kinds of wounds, as researchers continue to search for improved, cost-effective, and readily available agents to promote wound healing [9,12]. Although there is a sizeable number of reports that show mixed levels of effectiveness in the use of honey as a topical agent for the treatment of different types of ulcers, our search of numerous databases revealed a paucity of documentation on the use of honey in the treatment of ulcers in leprosy. The effectiveness of honey to promote the healing of ulcers in general remains uncertain. A Cochrane review has reported several studies with unclear outcomes for trials with honey on venous leg ulcers, diabetic foot ulcers, and mixed chronic wounds [9]. Most of the reported studies were adjudged by the reviewers to be of low quality due to imprecision and high risk of bias. The review suggested the high risk of bias in the reports were due to nonblinding of study participants and health care professionals. Statistical heterogeneity was evident across studies [9].

The current practice for the dressing of leprosy ulcers in Nigeria is the use of normal saline [13]. Honey is acknowledged to be safe for use in wound dressings [14], with only about 5% of patients reporting pain following dressing. There is also an undocumented concern of botulism disease due to infection with *Clostridium botulinum* [15] that is associated with the use of honey, as it is considered a suitable environment for the anaerobic bacteria to thrive. This study will evaluate the effectiveness of honey in the treatment of ulcers in leprosy in comparison with normal saline dressing.

Objectives

The aim of this study, Honey Experiment on Leprosy Ulcer (HELP), is to evaluate the healing properties of raw, unadulterated African honey in comparison with normal saline dressing for the treatment of leprosy ulcers.

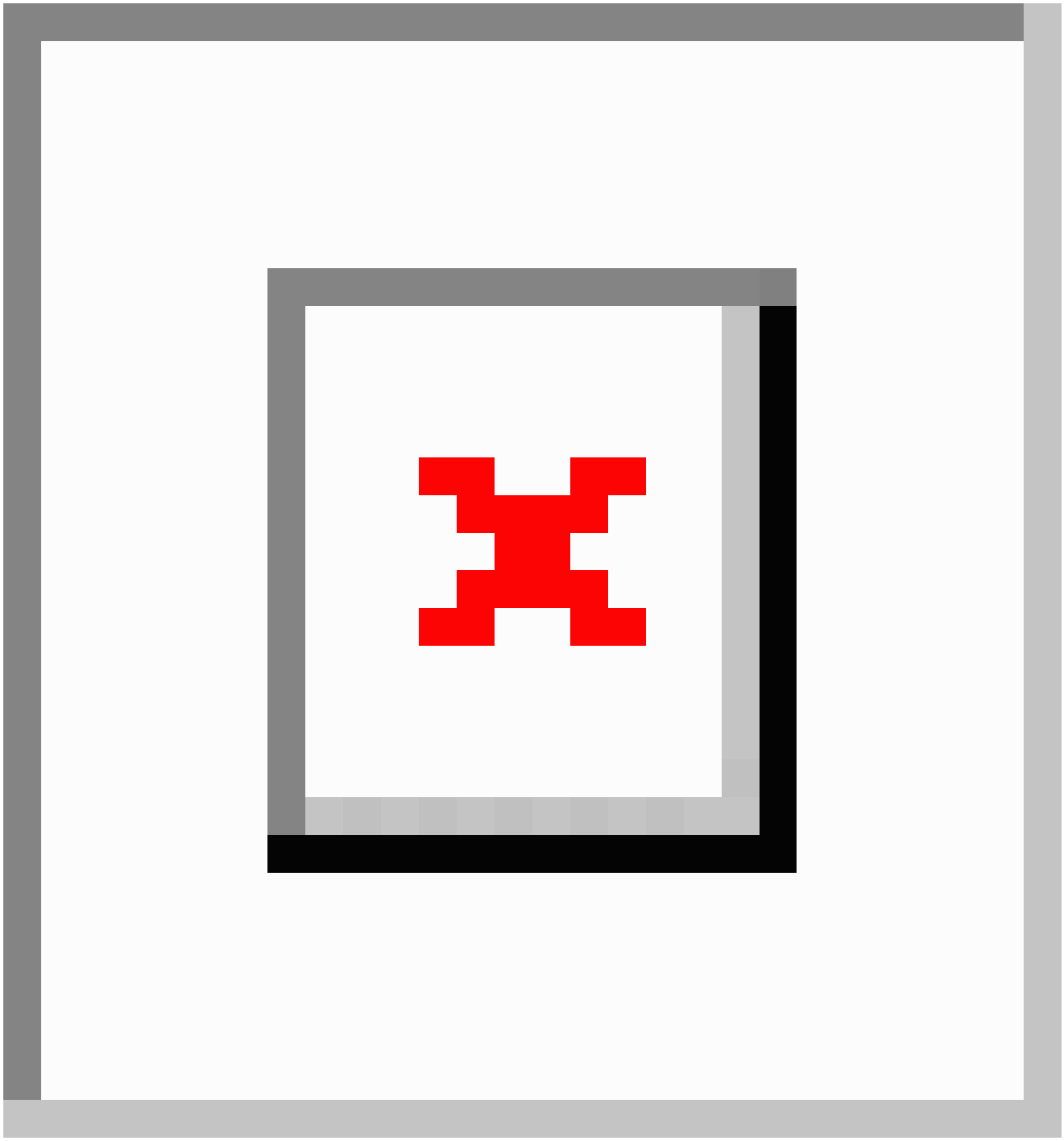
The study objectives are as follows:

1. Recruit 90 - 130 eligible, consenting people within 12 months.
2. Randomize the participants to receive ulcer dressing with either honey (intervention group) or normal saline (control group) twice a week.
3. Observe the rate of healing based on 2 observations per week (cm^2 per unit time, until either the ulcer has healed or 84 d have elapsed since randomization).
4. Observe the time to healing, defined as complete re-epithelization (up to a maximum of 84 d).
5. Monitor the rate of activity of study participants with pedometers.
6. Monitor the recurrence of treated ulcers or appearance of new ulcers and any anatomical changes in the limb at 6-month follow-up after randomization.

Trial Design

The study is a multicenter, prospective, single-blinded, parallel group, and 1:1 individually randomized controlled trial. The study duration is 48 months (maximum), with recruitment starting at month 5 and postdischarge follow-up starting at month 11. We aim to enroll between 90 and 130 participants over a 24-month recruitment period. The study pathway is shown in Figure 1.

Figure 1. Honey Experiment on Leprosy Ulcer (HELP) study pathway.



Methods

Study Setting

The study centers are The Leprosy Referral Hospital, Chanchaga in Minna, Niger State, North Central Nigeria, and St. Benedict Tuberculosis and Rehabilitation Hospital in Ogoja, Cross River State, South-South Nigeria. The main study will be preceded by prestudy due diligence.

The Leprosy Referral Hospital, Chanchaga in Minna, Nigeria, is a specialist hospital operated by the government of Niger State in collaboration with The Leprosy Mission (TLM) Nigeria. The hospital was established in 1940 and has 2 wards (eye and physiotherapy units), a theater, a laboratory, and a dispensary. The hospital is supported by an orthopedic workshop that is built and managed by TLM Nigeria, which produces assistive devices including wheelchairs, crutches, protective sandals, and artificial legs.

St. Benedict Tuberculosis and Rehabilitation Hospital is a tuberculosis and leprosy referral hospital located in Ogoja, a town in the northern part of Cross River State of Nigeria. It is owned by the Catholic Diocese of Ogoja, Cross River State, Nigeria. It provides these services in collaboration with the National Tuberculosis and Leprosy Control Programme and with the technical assistance of German Leprosy and TB Relief Association. The center has a community outreach program, which covers 31 communities in the Bekwarra, Ogoja, and Yala local government areas of Cross River State.

Ethical Considerations

The trial will be conducted in full conformance with the principles of the Declaration of Helsinki and Good Clinical Practice Guidelines. It will also comply with all applicable UK legislation and standard operating procedures for University of Birmingham–sponsored studies. Ethics approval has been granted in Nigeria for the study by the National Health Research Ethics Committee (approval FHREC/2022/01/09/04-02-22) and Niger State Ministry of Health (reference STA/495/Vol/199).

Eligibility Criteria

The local, medically qualified researchers employed on the grant will screen all ulcer admissions. The researchers will complete web-based eligibility forms for each patient with an ulcer. Consent will be sought by the research fellow from consecutive patients aged ≥18 years with a foot or leg ulcer between 2 and 20 cm^2 and not requiring surgery (eg, skin graft). The inclusion criteria are as follows:

1. Patients with a chronic foot ulcer of at least 6-week duration due to leprosy neuropathy.
2. Patients ≥18 years of age.
3. Ulcer with a surface area between 2 and 20 cm^2 inclusive.
4. Ulcer is clean, dry, and free from infection.
5. Patient can provide informed consent.

The exclusion criteria are as follows:

1. Any significant medical condition, laboratory abnormality, or psychiatric illness that would prevent the participants from participating in the study.
2. Ulcers with a surface area <2 or >20 cm^2.
3. Patient requires a skin graft.
4. Any condition that confounds the ability to interpret data from the study (ie, patients with HIV or tuberculosis under active treatment).
5. Any wound that has clinical microbial infections.
6. Diabetes or diabetic ulcer.
7. A patient has returned to the hospital having already taken part in the trial.

The wound dressing protocol for the HELP study is shown in Multimedia Appendix 1.

In the uncommon scenario where a participant has more than 1 foot ulcer, the largest ulcer will be selected as the index ulcer before randomization, and all ulcers will receive the same treatment (eg, if the participant is in the intervention group, all ulcers—not just the 1 being monitored for the trial—will be treated with honey). Observations will be made on all ulcers, but only the largest ulcer will be used in the primary analysis (see the *Discussion* section). Eligible patients will be offered entry in the trial at the point where their senior clinician judges them to be suitable for the honey treatment. This point arises once the lesion has been cleared of any debris or infection, typically by 7‐10 days after beginning treatment with or without antibiotic or debridement.

Swabs of the ulcer area will be taken to assess the wound for bacterial infections prior to the recruitment of participants into the study.

Informed Consent Collection

Researchers at the study center have been trained on Good Clinical Practice. They will be responsible for the screening of eligible participants and taking of consent. The participant information sheet has been translated into the local language (Hausa). All relevant information for the research participants on the study are contained in the participant information sheet. The information sheet will be given to the participants a day before enrollment into the study. The content of the information sheet will be explained to them verbally, and they will be encouraged to ask questions should they need more clarification. Written informed consent will be sought from the eligible participants after the study has been explained to them and they have taken time to decide to enroll into the study. The consent form will be signed or thumb- or fingerprinted by the participants before enrollment.

Additional Consent Provisions for the Collection and Use of Participant Data and Biological Specimens

The participant information sheet contains all information about the data to be collected and how the data will be stored and used. Biological specimens such as a swab of the ulcer surface will be obtained for the purpose of screening for bacterial infections prior to the recruitment of a participant into the study. Participants who refuse to give consent will continue to receive normal care from the hospital, but they will not be part of the study.

Participants can decide to withdraw their consent at any time during the study. Such participants will be given a form to complete, stating their reason for withdrawal. A participant may either withdraw fully or they may withdraw from treatment but consent to ongoing data collection.

Once a person consented to participate in the trial, baseline data will be collected. This will precede randomization.

Data will be collected directly onto electronic tablets, and the computer program (Research Electronic Data Capture [REDCap]; Vanderbilt University) will check the range of the information. Photographs of the ulcers will be captured using the tablet device. Each participant will be allocated a unique trial number, which will be included in all data entry forms and linked to the photographs. The level of activity (step count) will be recorded at each dressing change across both intervention and control groups until 84 days or discharge, whichever comes first. This date will enable us to check for postrandomization (“performance”) bias.

Interventions

Explanation for the Choice of Comparators

The participants in the control group will receive the usual care of twice weekly normal saline dressings only. Normal saline dressing is the standard ulcer dressing method currently in use in the hospitals where the trial is taking place.

Intervention Description

The intervention for this study is raw, unadulterated honey obtained from local bee farmers in North Central Nigeria. The honey samples have been examined at The National Institute

for Pharmaceutical Research and Development, Abuja, and confirmed to be free from microbial contaminants. The honey is stored in airtight plastic containers at room temperature and away from direct sunlight. The honey will be applied topically to the wound during dressing under strict hygienic conditions by using sterilized equipment.

The treatment will be applied at twice weekly dressing changes by local trained nurses or paramedics. These dressing changes are part of routine care and will thus be applied to the intervention and control groups. Dressing changes may take slightly longer for participants in the intervention group, but pain is unlikely as the affected limb has a loss of sensation. Participants in both groups have twice weekly dressing changes during their hospital stay until the ulcers are healed. Any missed sessions will be noted, but this will not be treated as a protocol deviation.

Interval Pilot to Monitor the Blinding of Observations

We are concerned that honey residue may be discernible by the observers of the ulcer outcome. We will therefore conduct an interval pilot to examine this issue. To ensure proper blinding of the trial, images of the ulcers from the second dressing change (following the honey or saline application at the first dressing change) for the first 10 enrolled participants will be sent to 3 independent assessors in Nepal. The assessors will look out for traces of honey residues that might interfere with blind assessment during the study. If the assessors are able to distinguish between cases treated with honey and controls, then we will consider alternating honey and saline dressings and making weekly observations rather than twice weekly observations.

We will also make a video recording of the first dressing change for the first 5 intervention and 2 control participants and send it to the 3 independent assessors. An uninterrupted video recording will be made from before the dressing is removed until after it has been replaced.

Relevant Concomitant Care Permitted or Prohibited During the Trial

In line with standard practice, participants will be discouraged from bearing weight on the ulcer site, since weight bearing and the patients' level of activity might affect the ulcer's rate of healing. Participants will be asked to wear pedometers on the nonaffected foot, and the level of activity will be recorded on the tablet at each dressing change from the first dressing change until 84 days from randomization or discharge, whichever comes first.

Provisions for Posttrial Care

The date of discharge will be noted, along with participant contact details, addresses, and contact details for at least 1 family member. The participants will be routinely contacted 6 months after randomization. The treated ulcer area will be examined and photographed. The researcher will take photographs of the ulcer area (healed in the majority of cases). The dates covering any readmissions to any hospital for the treatment of the "trial ulcer" will be recorded.

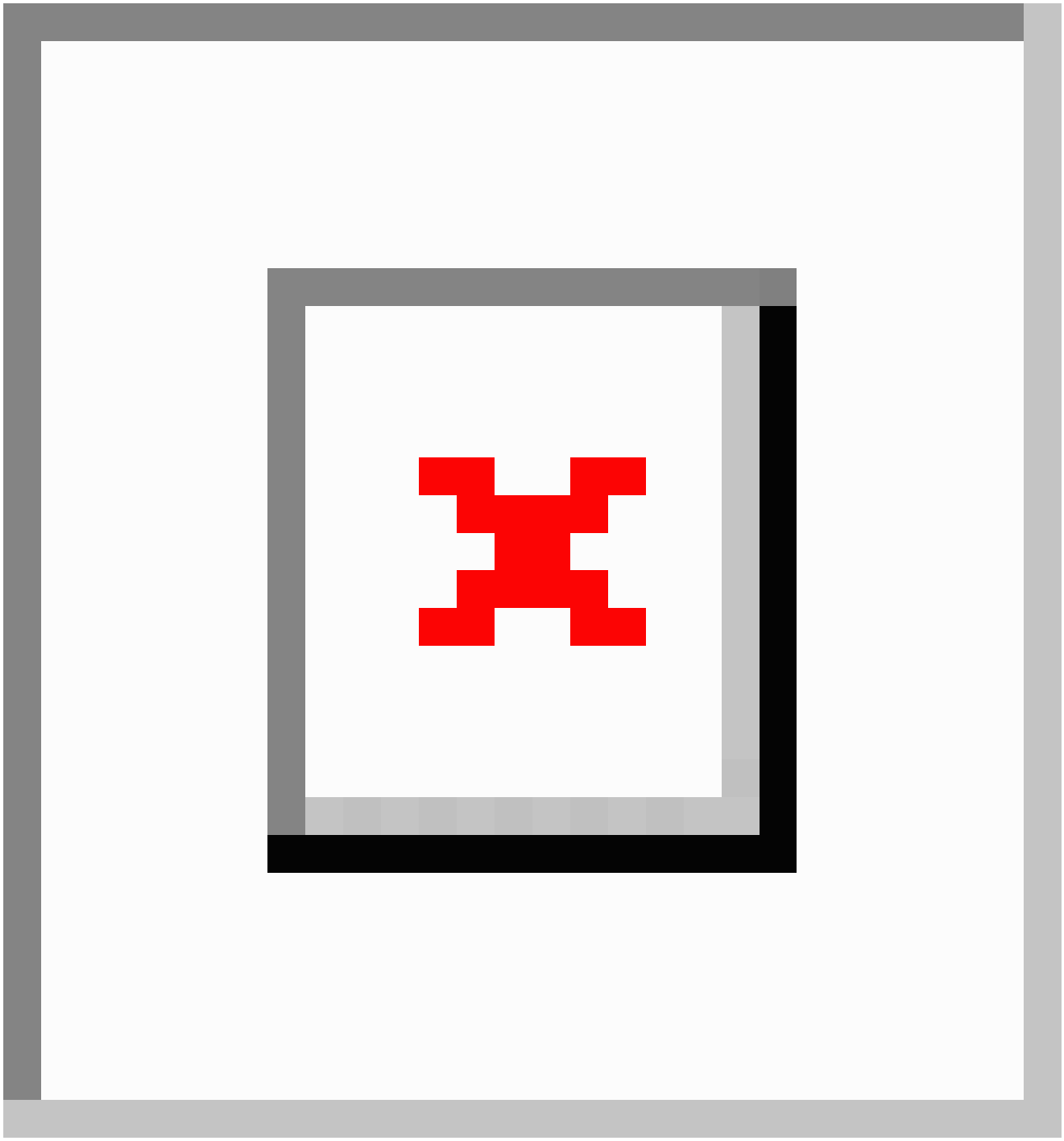
Outcomes

We define 2 main outcomes relating to ulcer healing. The main outcomes will be (1) rate of healing based on 2 observations per week (until healed or 84 d from randomization) and (2) time to healing defined as complete re-epithelization (up to 84 d). Secondary outcomes will be (1) the recurrence of treated ulcer and (2) appearance of a new ulcer. Long-term outcomes will be measured at the follow-up 6 months after randomization, which will be (1) the number of days hospitalized prior to discharge and in total (to include any readmission related to leprosy ulcers) by 6 months and (2) the number of visits to any health care facility from discharge to the end of follow-up at 6 months.

Participant Timeline

Participants' timeline through the trial is outlined in Figure 2.

Figure 2. Timeline (in months) for the Honey Experiment on Leprosy Ulcer (HELP) trial (based on the most pessimistic estimate).



Sample Size

Recruitment will continue for 24 months. We expect to enroll between 90 and 130 participants. Sample size was based on the 2 clinical outcomes: the rate of healing and time to complete re-epithelialization. We expect at least 70% of ulcers to heal within 84 days with standard care (according to a recent study of neuropathic ulcers, with over half due to leprosy [16]). If we assume that the intervention will increase this proportion to 90% and hazards are constant and proportional (so that the hazard ratio is 1.91 for discharge), for a 2-sided test of the hazard ratio with a type I error of 5%, a statistical power of 80%, and a 1:1 allocation ratio, 47 individuals are required in each group. With 130 participants, this allows for a dropout rate of up to 40% to achieve a power of 80%. At the most pessimistic sample size of 90, with a dropout rate of 40%, our minimum detectable effect size (ie, the effect size that achieves a power of 80% with 33 patients per arm) is a hazard ratio of 2.15, or 92.5% of patients in the treatment group being discharged by the end of the trial period. At the most optimistic sample size of 130 with no dropout, our minimum detectable effect size is a hazard ratio of 1.74, or 87.3% of patients being discharged in the treatment group. All calculations are based on a log-rank test.

We took a conservative approach and based the sample size calculation on the healing outcome and model with lower efficiency to ensure an adequate sample size for all main outcomes, since our inferential approach is not based on

statistical significance but on a consideration of effect sizes and a comparison and triangulation of the totality of evidence. We are not particularly concerned with being "overpowered" or with a problem of multiple comparisons; instead, we will focus on ensuring a sufficient sample size to estimate clinical effectiveness to a reasonable degree of precision [17,18].

Recruitment

Eligible individuals are identified by the on-site clinical research team and are invited to take part in the study.

Assignment of Interventions: Allocation

Sequence Generation

Participants will be enrolled sequentially, stratified by ulcer size (above or below 10 cm^2) and randomly allocated (1:1) to undergo honey treatment or usual care with normal saline using a "digital sealed envelope" method [19].

Concealment Mechanism

An allocation table will be generated remotely at the trials office at The University of Ibadan. A permuted block randomization method will be used to generate the randomization sequence within each stratum. Randomly selecting blocks of size 2, 4, 6, or 8 will be generated in order to maintain balance between the numbers allocated to each of the 2 groups and to ensure allocation concealment. The generated table will be uploaded into the REDCap database platform to be used for participant enrollment. Access to the allocation table will be restricted. Trial staff in Nigeria will not have access to the allocation table. When a participant's details are submitted, the trial arm and a unique study number will be assigned and revealed to the local clinician so that the randomized group that the participant is assigned to cannot be altered.

Implementation

A trial statistician at The University of Ibadan will generate the allocation sequence while researchers and clinical staff on-site will enroll and assign participants to the respective control or intervention groups.

Assignment of Interventions: Blinding

Researcher Blinding

Only the overall on-site research supervisor, the database managers at the University of Birmingham and the University of Ibadan Trials Unit, the clinical staff carrying out dressing changes in the room designated for this purpose, and the participants themselves will be aware of the participants' randomly assigned group. Ward staff will not be informed. The assessors in Nepal will be blinded from the treatment provided for the randomly assigned groups.

Procedure for Unblinding if Needed

There is no provision for the unblinding of trial participants during the trial.

Data Collection and Management

Plans for the Assessment and Collection of Outcomes

Standardized photographs [20] will be taken twice weekly of the dressing changes during the in-patient stay for participants in both intervention and control groups. Any residual honey will be gently removed with a wet swab. Photographs will be taken using the built-in camera in the tablet devices (Samsung Galaxy Tab S7, 13Mp). The photographs will be taken perpendicular to the ulcer. For calibration purposes, a 3-cm clean paper ruler with the date and participant's trial identification number will be placed in the photograph frame above or below the ulcer but at the level of the skin. The photographs will then be uploaded onto the REDCap database platform and accessed by database managers at the University of Birmingham. The ulcer photos will then be randomized and sent to the blinded assessors in Nepal for the measurement of ulcer dimensions. The photographs will be evaluated digitally by the designated observer in Nepal using the PictZar Digital Planimetry Software with the electronic PUSH Tool (National Pressure Injury Advisory Panel [21]). The observer will delineate an area of interest by manually "painting" the ulcer area with color using a computer mouse. The software will then calculate the ulcer dimensions based on this profile.

The assessors will be blinded to the treatment allocated to the participant. The assessors are all experienced in leprosy, ulcer care, and the measurement of ulcers using the PictZar software. All photographs from a given participant will be assigned to the observers at random. To ensure that the measurements are not all relegated to the end of the study, they will be made in batches of 10 participants—when they reach the completion of their treatments. A proportion (20/100, 20%) of all ulcer photographs will be measured by 2 assessors to estimate interrater reliability. These photographs will be selected at random.

Treatment will stop when the local clinician determines that complete epithelization has taken place. Photographs will be taken at this point and at the follow-up visit.

Plans to Promote Participant Retention and Complete Follow-Up

TLM Nigeria has a meal subsidy program for all in-patients at The Leprosy Referral Hospital, Chanchaga. The meal subsidy program will continue through the period of this study. It is expected that the meal subsidy program will be sufficient to keep the patient at the hospital from the period of admission up to the time of discharge. Phone contacts of participants or their close relatives will be collected and used for postdischarge follow-up.

Data Management

All data generated from this study will be classified according to the University of Birmingham Information Security Framework. All data will be collected and stored electronically to reduce data entry errors, such as contradicting answers. Data will be reported on an electronic case report form, and all local staff will be trained to collect data directly onto electronic tablets. Data will be acquired and stored on the REDCap

database platform, with access restricted by passwords at both the University of Ibadan and the local site in Nigeria. Each participant will be allocated a unique study number when they agree to participate (and before randomization), which will be used on all documents. A master list linking a trial participant number to their identity (name) will be retained by the hospital securely in a locked filing cabinet. This is necessary so that the notes of any person who withdraws consent for data storage can be removed. The trial participant master list will be stored separately from patient lists and the trial data.

Confidentiality

Participant confidentiality will be ensured throughout the study and no participant-identifying information will be released to anyone outside the project. Confidentiality will be secured through several mechanisms. Each participant will be assigned a study subject ID, which will be used on all study forms. Any study forms and paper records that contain personal identifier information (eg, address lists and phone lists) will be kept secured and locked at the trial sites. Access to all participant data and information at the sites will be restricted to authorized personnel.

Participants will not be identified by name in any reports or publications, nor will the data be presented in such a way that the identity of individual participants could be inferred. Analysis files created for further study by the scientific community will have no participant identifiers. The trial data will be kept under lock for up to 10 years, after which they will be destroyed. Only authorized persons will be granted access to the trial data.

Plans for Collection, Laboratory Evaluation, and Storage of Biological Specimens for Genetic or Molecular Analysis in This Trial or for Future Use

At the pre-enrollment stage, we will collect a swab of the ulcer surface to examine the wound for bacterial infections. Afterward, we will not collect or retain biological specimens for any future use in this trial.

Statistical Methods

Statistical Methods for Primary and Secondary Outcomes

Time to healing will be analyzed using a Cox proportional hazards model with and without adjustment for baseline characteristics (trial ulcer area and participants' age), allowing for right censoring. For the rate of healing, we will define the outcome ulcer size in cm^2 at each time point and include in the model time since admission, treatment status, and their interaction. We will analyze this model using a linear mixed-effects model with participant-level random effects and both with and without adjustment for participant characteristics. Given there are multiple clinical outcomes (2 outcomes, with and without adjustment), we will adjust reported *P* values for multiple testing using a step-down method, which provides an efficient means of controlling the family-wise error rate [22]. We will derive the approximate distributions of the test statistics to perform the step-down procedure using a permutation test approach, by simulating 10,000 rerandomizations of the individuals [23].

Interim Analyses

An interim analysis will be conducted when at least 49 (three-eighths [37.5%] of the 130 target) participants have been followed up for of 84 days or discharged, whichever comes first. The rationale for this analysis is the detection of a "penicillin-like" benefit or statistically significant negative effect of the treatment on either primary outcome. The statistical methods will be as specified above. A statistical threshold of 0.01, one-sided (0.02, two-sided), will be used for either primary outcome. If either threshold is exceeded, or if the Independent Data Monitoring Committee (IDMC) has other concerns, then the Trial Steering Committee (TSC) will be advised accordingly. If the IDMC wishes to meet again before trial conclusion, they will be able to do so. Only the trial statistician and the IDMC will see the "unblinded" data, unless the TSC is informed.

Methods for Additional Analyses

We will also compare average daily step count between treatment and control groups as a simple difference in means (1-tailed *t* test). Since 1 group may stay longer in hospital than the other and since there may be an interaction between the rate of healing and step count, we will compare step counts over preset periods at 7, 14, and 42 days.

Analytical Methods to Handle Protocol Nonadherence and Statistical Methods to Handle Missing Data

In the unlikely event of missing data by any reason, such as the withdrawal of participant before the completion of treatment, and that the rate differs between groups, a sensitivity analysis will be carried out. The pattern and rates of missing data that are particularly related to the end points will be explored. Strategies such as multiple imputation within the sensitivity analysis will be explored instead of imputing missing values.

Plans to Give Access to the Full Protocol, Participant-Level Data, and Statistical Code

The full protocol, nonidentifiable participant-level data, and statistical code may be available for sharing once the trial has ended. All requests will be approved by Chief Investigator RL.

Oversight and Monitoring

Composition of the Coordinating Center and TSC

TLM Nigeria is the study sponsor and will oversee the study process in Nigeria. The University of Ibadan will provide data management services and perform quality checks, whereas the University of Birmingham will house the data securely and perform quality checks.

The Trial Management Group (TMG) includes individuals at the University of Birmingham; The Leprosy Referral Hospital, Chanchaga; TLM Nigeria; the German Leprosy and TB Relief Association/RedAid Nigeria; and the University of Ibadan Clinical Trials Unit who are responsible for the day-to-day management of the trial. The TMG will meet monthly by teleconference, but this may be more frequent if deemed necessary by the members.

The TSC will provide overall supervision of the trial and ensure that it is being conducted in accordance with the principles of

Good Clinical Practice and other relevant regulations. The TSC will have oversight of the trial. It will send its report to the overall Research and Innovation for Global Health Transformation program steering committee, but the TSC will have the final say with the running of the trial itself. The TSC will meet either face-to-face or via teleconferencing. Meetings will be scheduled for before the enrollment of participants begins, following each meeting of the IDMC, and then during the analysis phase or more often if required.

The IDMC's Composition, Role, and Reporting Structure

The IDMC consists of individuals who have no conflict of interest within the study. Safety and efficacy data will be supplied, in strict confidence, for review by the IDMC after the 49th participant has been followed up for of 84 days or discharged, whichever comes first. The IDMC will be asked to give advice on whether the accumulated data from the trial, together with the results from other relevant research, justifies the continuing recruitment of further participants. The IDMC will meet either face-to-face or by teleconferencing. An emergency meeting may also be convened if a safety issue is identified. The IDMC will escalate any issues and recommendations to the TSC, who will then make decisions based on these recommendations.

Adverse Event Reporting and Harms

Overview

The principal investigator in Nigeria will be responsible for recording all adverse events (AEs) and reporting any serious AEs to the University of Birmingham and the University of Ibadan within 24 hours of the research staff becoming aware of the event. A serious AE form will be available on the tablets used to collect data, and we will maintain a database of all safety events or AEs. The forms will be reviewed by the TMG, which meets monthly, and by the project manager, if required. The TSG will periodically review all safety data and liaise with the IDMC regarding any safety issues.

Any deaths will be reported to the sponsor, irrespective of whether the death is related to the disease progression, the intervention, or an unrelated event. Only deaths that could plausibly be caused by the intervention will be reported to the sponsor immediately.

Frequency and Plans for Auditing Trial Conduct

The trial is audited and monitored by the sponsor, TLM Nigeria.

Plans for Communicating Important Protocol Amendments to Relevant Parties

Any protocol amendment will be reported to the TMG to approve the change. The amendment will be sent to the sponsor to confirm substantiality and then to the National Health Research Ethics Committee for approval.

Dissemination Plans

The results of the trial will first be reported to trial collaborators. The main report will be drafted by the trial coordinating team, and the final version will be agreed by the TSC before submission for publication, on behalf of the collaboration.

We will present our work at conferences such as the annual conference of the Neglected Tropical Disease NGO Network, of which TLM is a member; the Health Systems Global Conference; and the International Leprosy Congress in 2024. Other dissemination plans include bite-sized research reports in lay format, public announcements in communities in low- and middle- income countries, policy briefings, print and web-based media, the chief investigator's news blog (680+ subscribers), and institutional and professional social media accounts and websites.

Results

The recruitment of trial participants began on March 14, 2022, and has been continuing for approximately 24 months.

Discussion

This study protocol describes a clinical trial with honey as an intervention for the dressing of leprosy ulcers. It is a single-center, comparative, prospective, single-blinded, parallel-group, and blocked-stratified randomized controlled trial.

Leprosy is regarded as a neglected tropical disease because of its near absence from the global health agenda [24]; as such, very little attention has been given to the treatment of ulcers in leprosy in spite of its debilitating nature and the social stigma it attracts. Methods such as laser therapy [25], zinc tape [26], pentoxifylline injection [27], amniotic membrane gel [28], phenytoin suspension [29], and leukocyte- and platelet-rich fibrin [3] have been tested on leprosy foot ulcers as the search for the most suitable treatment continues. Normal saline has been the most common comparator in most of the studies on ulcers in leprosy [30]. Among the challenge with the normal saline dressing are the slow rate of healing, the possibility of ulcers getting infected with pathogens [31], and the likelihood of the recurrence of treated ulcers. A Cochrane review [30] noted that previously published evidence is limited, due to a high or unclear risk of bias (selection, performance, detection, or attrition), imprecision due to the small number of participants, indirectness due to poor outcome measures, and inapplicable interventions.

Honey is considered as a cheaper and more readily accessible alternative treatment for many types of wounds. Although honey has been known for centuries to promote wound healing, there are only a few controlled clinical trials that assess its efficacy. While honey is noted to have wound healing properties, including debridement, deodorization, and antimicrobial properties [14], its use may result in painful sensation in about 5% of persons, although this is unlikely to occur in leprosy and diabetes [15]. The hypothetical risk of botulism will be mitigated by regularly screening the honey sample. Detailed procedures for the study have been described in this protocol. We expect that the outcome of this study will provide valuable evidence on the use of honey for the treatment of ulcers in leprosy.

Acknowledgments

We acknowledge the contribution of the Public Health Department of the Niger State Ministry of Health and The Leprosy Referral Hospital, Chanchaga to the clinical trial. We also acknowledge members of the communities affected with leprosy in Nigeria for their assurance of cooperation during pretrial engagements. We thank the chair and members of the Trial Steering Committee—Dr Doyin Odubanjo, Professor Kara Hanson, Dr Paul Saunderson, and Dr Jerry Joshua—for their constructive advice on the trial. We also acknowledge the support of Dr Indra Napit, Dilip Shrestha, Karuna Neupane, and Anjali Shrivastav from The Leprosy Mission Nepal for their support in training for the trial. This research was funded by the National Institute for Health Research (NIHR: 200132) using UK aid from the UK government to support global health research. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Data Availability

The full protocol, nonidentifiable participant-level data, and statistical code may be available for sharing once the trial has ended. All requests will be approved by Chief Investigator RL (r.j.lilford@bham.ac.uk).

Authors' Contributions

POS, PAT, IOO, AM, and LCU are local coinvestigators and have contributed to the study conception, design, and intellectual contents of the protocol. SMC is the research project manager at the University of Birmingham and has contributed to the study conception, design, and intellectual contents of the protocol. JS is the study senior project manager at the University of Birmingham and has contributed to the study conception, design, and intellectual content of the protocol. OI is research fellow in the study and has contributed to the study design and intellectual content of the protocol. SW and JA contributed to the design of the evaluation, sample size calculation, randomization, and quantitative sample selection and critically evaluated the intellectual content of the protocol. MA and A Oladejo are the project managers at the University of Ibadan and critically evaluated the intellectual content of the protocol. JA contributed to the design of the trial randomization and intellectual content of the protocol. A Omigbodun leads the clinical trials team at the University of Ibadan and contributed to the study design and intellectual content of the protocol. SU is the local chief investigator and contributed to the study conception and intellectual content of the protocol. RL is the principal investigator at the University of Birmingham, contributed to the study conception and design, and critically evaluated the intellectual content of the protocol. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Wound dressing protocol.
[DOCX File, 30 KB - xmed_v5i1e50970_app1.docx]

References

1. Miyashiro D, Cardona C, Valente NYS, Avancini J, Benard G, Trindade MAB. Ulcers in leprosy patients, an unrecognized clinical manifestation: a report of 8 cases. BMC Infect Dis 2019 Nov 29;19(1):1013. [doi: 10.1186/s12879-019-4639-2] [Medline: 31783808]
2. Towards zero leprosy. global leprosy (Hansen's disease) strategy 2021-2030. World Health Organization. 2021 Apr 21. URL: https://www.who.int/publications/i/item/9789290228509 [accessed 2024-02-09]
3. World Health Organization (Organisation mondiale de la Santé). Weekly epidemiological record (relevé épidémiologique hebdomadaire). Wkly Epidemiol Rec 2020;95(36):417-440 [FREE Full text]
4. Leprosy. World Health Organization. 2023 Jan 27. URL: https://www.who.int/news-room/fact-sheets/detail/leprosy [accessed 2024-02-09]
5. Napit IB, Shrestha D, Bishop J, et al. An individual randomised efficacy trial of autologous blood products, leukocyte and platelet-rich fibrin (L-PRF), to promote ulcer healing in leprosy in Nepal: the TABLE trial protocol. Trials 2021 Jul 15;22(1):453. [doi: 10.1186/s13063-021-05392-5] [Medline: 34266456]
6. Riyaz N, Sehgal VN. Leprosy: trophic skin ulcers. Skinmed 2017 Feb 1;15(1):45-51. [Medline: 28270310]
7. Okeniyi JAO, Olubanjo OO, Ogunlesi TA, Oyelami OA. Comparison of healing of incised abscess wounds with honey and EUSOL dressing. J Altern Complement Med 2005 Jun;11(3):511-513. [doi: 10.1089/acm.2005.11.511] [Medline: 15992237]
8. Sato T, Miyata G. The nutraceutical benefit, part III: honey. Nutrition 2000 Jun;16(6):468-469. [doi: 10.1016/s0899-9007(00)00271-9] [Medline: 10869910]

9. Jull AB, Cullum N, Dumville JC, Westby MJ, Deshpande S, Walker N. Honey as a topical treatment for wounds. Cochrane Database Syst Rev 2015 Mar 6;2015(3):CD005083. [doi: 10.1002/14651858.CD005083.pub4] [Medline: 25742878]
10. Mphande ANG, Killowe C, Phalira S, Jones HW, Harrison WJ. Effects of honey and sugar dressings on wound healing. J Wound Care 2007 Jul;16(7):317-319. [doi: 10.12968/jowc.2007.16.7.27053] [Medline: 17708384]
11. Subrahmanyam M, Ugane SP. Honey dressing beneficial in treatment of Fournier's gangrene. Indian J Surg 2004;66(2):75-77 [FREE Full text]
12. Al-Waili NS, Saloom KY. Effects of topical honey on post-operative wound infections due to gram positive and gram negative bacteria following Caesarean sections and hysterectomies. Eur J Med Res 1999 Mar 26;4(3):126-130. [Medline: 10085281]
13. Lim JK, Saliba L, Smith MJ, McTavish J, Raine C, Curtin P. Normal saline wound dressing--is it really normal? Br J Plast Surg 2000 Jan;53(1):42-45. [doi: 10.1054/bjps.1999.3246] [Medline: 10657448]
14. Tashkandi H. Honey in wound healing: an updated review. Open Life Sci 2021 Oct 6;16(1):1091-1100. [doi: 10.1515/biol-2021-0084] [Medline: 34708153]
15. Simon A, Traynor K, Santos K, Blaser G, Bode U, Molan P. Medical honey for wound care—still the 'latest resort'? Evid Based Complement Alternat Med 2009 Jun;6(2):165-173. [doi: 10.1093/ecam/nem175] [Medline: 18955301]
16. Rai S, Gupta AK, Kumar D, Sharma VP, Agarwal AK. Prospective analytical study of assessment of off loading by total contact cast in treatment of non healing plantar ulcers in anaesthetic foot. Lepr Rev 2016 Mar;87(1):71-77. [doi: 10.47276/lr.87.1.71] [Medline: 27255060]
17. Gelman A, Vehtari A, Simpson D, et al. Bayesian workflow. arXiv. Preprint posted online on Nov 3, 2020. [doi: 10.48550/arXiv.2011.01808]
18. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. J Res Educ Eff 2012 Apr 3;5(2):189-211. [doi: 10.1080/19345747.2011.618213]
19. Lim CY, In J. Randomization in clinical studies. Korean J Anesthesiol 2019 Jun;72(3):221-232. [doi: 10.4097/kja.19049] [Medline: 30929415]
20. Bowen AC, Burns K, Tong SYC, et al. Standardising and assessing digital images for use in clinical trials: a practical, reproducible method that blinds the assessor to treatment allocation. PLoS One 2014 Nov 6;9(11):e110395. [doi: 10.1371/journal.pone.0110395] [Medline: 25375169]
21. Pressure Ulcer Scale for Healing (PUSH). National Pressure Injury Advisory Panel. URL: https://npiap.com/page/PUSHTool [accessed 2024-02-09]
22. Romano JP, Wolf M. Exact and approximate stepdown methods for multiple hypothesis testing. J Am Stat Assoc 2005;100(469):94-108. [doi: 10.1198/016214504000000539]
23. Thompson J, Davey C, Hayes R, Hargreaves J, Fielding K. Swpermute: permutation tests for stepped-wedge cluster-randomised trials. Stata J 2019 Dec 1;19(4):803-819. [doi: 10.1177/1536867X19893624] [Medline: 32565746]
24. Neglected tropical diseases. World Health Organization. 2024 Jan 9. URL: https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases [accessed 2024-02-09]
25. The use of low level laser therapy for wound healing in leprosy patients. ClinicalTrials.gov. 2009 Dec 15. URL: https://clinicaltrials.gov/study/NCT00860717 [accessed 2024-02-09]
26. Kumar A, Lakshmanan M. Adhesive zinc tape treatment of uncomplicated ulcers amongst leprosy outpatients. Lepr Rev 1986 Mar;57(1):45-51. [doi: 10.5935/0305-7518.19860008] [Medline: 3702580]
27. Jull A, Waters J, Arroll B. Pentoxifylline for treatment of venous leg ulcers: a systematic review. Lancet 2002 May 4;359(9317):1550-1554. [doi: 10.1016/S0140-6736(02)08513-6] [Medline: 12047963]
28. Natallya FR, Herwanto N, Prakoeswa C, Indramaya DM, Rantam FA. Effective healing of leprosy chronic plantar ulcers by application of human amniotic membrane stem cell secretome gel. Indian J Dermatol 2019;64(3):250. [doi: 10.4103/ijd.IJD_6_17] [Medline: 31148869]
29. Bhatia A, Nanda S, Gupta U, Gupta S, Reddy BSN. Topical phenytoin suspension and normal saline in the treatment of leprosy trophic ulcers: a randomized, double-blind, comparative study. J Dermatolog Treat 2004 Sep;15(5):321-327. [doi: 10.1080/09546630410018085] [Medline: 15370401]
30. Reinar LM, Forsetlund L, Lehman LF, Brurberg KG. Interventions for ulceration and other skin changes caused by nerve damage in leprosy. Cochrane Database Syst Rev 2019 Jul 31;7(7):CD012235. [doi: 10.1002/14651858.CD012235.pub2] [Medline: 31425632]
31. Bowler PG, Duerden BI, Armstrong DG. Wound microbiology and associated approaches to wound management. Clin Microbiol Rev 2001 Apr;14(2):244-269. [doi: 10.1128/CMR.14.2.244-269.2001] [Medline: 11292638]

Abbreviations

AE: adverse event
HELP: Honey Experiment on Leprosy Ulcer
IDMC: Independent Data Monitoring Committee
REDCap: Research Electronic Data Capture

TLM: The Leprosy Mission
TMG: Trial Management Group
TSC: Trial Steering Committee

Edited by E Meinert; submitted 18.07.23; peer-reviewed by Anonymous, Anonymous; revised version received 27.12.23; accepted 13.01.24; published 01.03.24.

Please cite as:
Udo S, Ogbu Sunday P, Tsaku PA, Oladejo IO, Meka A, Ugwu LC, Ajisola M, Akinyemi J, Oladejo A, Omigbodun A, Choudhury SM, Sartori J, Ilozumba O, Watson S, Lilford R
Raw, Unadulterated African Honey for Ulcer Healing in Leprosy: Protocol for the Honey Experiment on Leprosy Ulcer (HELP) Randomized Controlled Trial
JMIRx Med 2024;5:e50970
URL: https://xmed.jmir.org/2024/1/e50970
doi:10.2196/50970

Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study

Theodore Lytras, MPH, MD, PhD

School of Medicine, European University Cyprus, Engomi, Cyprus

Corresponding Author:
Theodore Lytras, MPH, MD, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/43341

Companion article: https://med.jmirx.org/2024/1/e59639

Companion article: https://med.jmirx.org/2024/1/e59638

Companion article: https://med.jmirx.org/2024/1/e59637

Abstract

Background: Our previous analysis showed how in-hospital mortality of intubated patients with COVID-19 in Greece is adversely affected by patient load and regional disparities.

Objective: We aimed to update this analysis to include the large Delta and Omicron waves that affected Greece during 2021-2022, while also considering the effect of vaccination on in-hospital mortality.

Methods: Anonymized surveillance data were analyzed from all patients with COVID-19 in Greece intubated between September 1, 2020, and April 4, 2022, and followed up until May 17, 2022. Time-split Poisson regression was used to estimate the hazard of dying as a function of fixed and time-varying covariates: the daily total count of intubated patients with COVID-19 in Greece, age, sex, COVID-19 vaccination status, region of the hospital (Attica, Thessaloniki, or rest of Greece), being in an intensive care unit, and an indicator for the period from September 1, 2021.

Results: A total of 14,011 intubated patients with COVID-19 were analyzed, of whom 10,466 (74.7%) died. Mortality was significantly higher with a load of 400-499 intubated patients, with an adjusted hazard ratio (HR) of 1.22 (95% CI 1.09-1.38), rising progressively up to 1.48 (95% CI 1.31-1.69) for a load of ≥800 patients. Hospitalization away from the Attica region was also independently associated with increased mortality (Thessaloniki: HR 1.22, 95% CI 1.13-1.32; rest of Greece: HR 1.64, 95% CI 1.54-1.75), as was hospitalization after September 1, 2021 (HR 1.21, 95% CI 1.09-1.36). COVID-19 vaccination did not affect the mortality of these already severely ill patients, the majority of whom (11,944/14,011, 85.2%) were unvaccinated.

Conclusions: Our results confirm that in-hospital mortality of severely ill patients with COVID-19 is adversely affected by high patient load and regional disparities, and point to a further significant deterioration after September 1, 2021, especially away from Attica and Thessaloniki. This highlights the need for urgent strengthening of health care services in Greece, ensuring equitable and high-quality care for all.

(JMIRx Med 2024;5:e43341) doi:10.2196/43341

KEYWORDS

COVID-19; pandemic; health care disparities; intensive care unit; right to health; quality of care; intubation; mortality; health disparity; health inequality; surveillance data; inpatient; mortality; COVID-19 patient; hospitalization; ICU; disparity; inequality; surveillance; health care system; Greece; region; Delta; Omicron; vaccination; vaccine; public health; patient load; deterioration; time

Introduction

During the COVID-19 pandemic, an association between high patient load and in-hospital mortality has been identified in different settings [1,2]. We previously showed how the mortality of intubated patients with COVID-19 in Greece is affected by regional disparities and patient load, even without exceeding capacity [3]; that analysis did not consider COVID-19 vaccination and only covered the period until May 2021, thereby missing the large Delta and Omicron variant pandemic waves that followed, which were accompanied by a large number of deaths.

In this context, we aimed to update our analysis in order to (1) validate our previously published findings regarding the association between patient load, regional disparities, and mortality of intubated patients with COVID-19; (2) examine whether any changes in this association happened during the recent period, when Delta and Omicron were in circulation; and (3) examine whether vaccination reduces the mortality of already severely ill patients with COVID-19.

Methods

Overview

Our methods have been described before [3]; our study followed the retrospective cohort design. Briefly, we obtained anonymized patient data from the Greek National Public Health Organization (NPHO) for all cases intubated between September 1, 2020, and April 3, 2022, including vaccination status (number of doses received); dates of intubation, extubation, intensive care unit (ICU) admission, and discharge; and outcome at discharge (alive or dead); we followed these cases up to May 17, 2022.

Follow-up time between intubation and extubation or death was split finely into days, and Poisson regression was used to estimate the hazard of dying as a function of fixed and time-varying covariates [4]. Deaths occurring up to 5 days after extubation were classified as deaths at the end of follow-up. We included in the regression the following covariates: the daily total of intubated patients with COVID-19 as an indicator of health care system stress, age (as a natural cubic spline with 1 internal knot), sex, vaccination status (as a categorical variable), a linear time trend, ICU hospitalization (vs non-ICU), hospital region (the metropolitan regions of Attica and Thessaloniki vs the rest of Greece), an indicator for the period from September 1, 2021, and an interaction between period and hospital region. Model-based effect estimates were used to calculate population-attributable fractions [5] for the different covariates included.

Ethical Considerations

No ethical approval was necessary as we used only anonymous surveillance data from which no patient could be identified. These data were provided to the author by the NPHO for epidemiological analysis purposes in the context of the COVID-19 pandemic response.

Results

The distribution over time of new intubations, deaths among the study population, and total intubated patients with COVID-19 is illustrated in Figure 1. From August 2021, intubations and deaths gradually increased concurrently with the increased circulation of the Delta variant, especially so from November 2021. Then, after January 2022, circulation of the Omicron variant prolonged this epidemic wave further, with gradual de-escalation until the end of the study period.

A total of 14,011 intubated COVID-19 patients were analyzed, of whom 10,466 (74.7%) died and 11,944 (85.2%) were unvaccinated (Table 1). Most patients spent part or all of their hospital stay in an ICU (12,902/14,011 patients, representing 239,201/250,978 person-days in total). Among those not admitted in an ICU, nearly all died (1084/1109 patients, 97.7%), compared to 72.7% (9382/12,902 patients) among those admitted (P<.001).

Figure 1. Distribution over time of total intubated patients with COVID-19, new intubations, and deaths among the study population in Greece from September 1, 2020, to May 17, 2022.

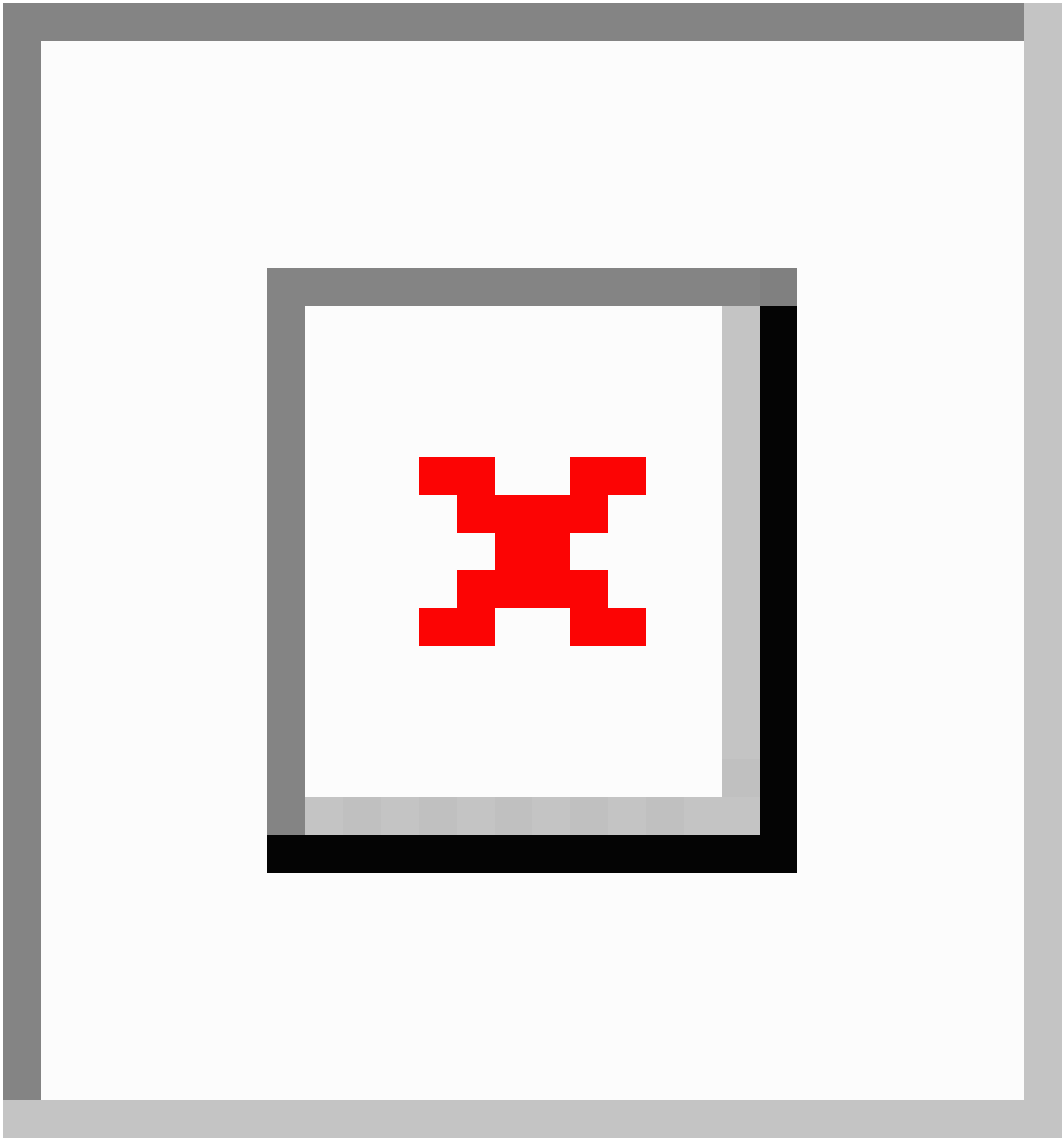


Table . Characteristics of the study population: patients with laboratory-confirmed COVID-19 in Greece intubated between September 1, 2020, and April 3, 2022.

Characteristics		Total participants (N=14,011)	Surviving participants (n=3545)[a]	Deceased participants (n=10,466)[a]	*P* value[b]
Age (years), median (IQR)		68 (59-75)	60 (52-68)	70 (62-77)	<.001
Sex, n (%)					.03
	Female	5312 (100)	1290 (24)	4022 (76)	
	Male	8699 (100)	2255 (26)	6444 (74)	
Vaccination status, n (%)					<.001
	Unvaccinated	11,944 (100)	3157 (26)	8787 (74)	
	Vaccinated with 1 dose	574 (100)	127 (22)	447 (78)	
	Vaccinated with 2 doses	1054 (100)	175 (17)	879 (83)	
	Vaccinated with 3 doses	439 (100)	86 (20)	353 (80)	
Hospitalization type, n (%)					<.001
	In the ICU[c]	12,902 (100)	3520 (27)	9382 (73)	
	Outside the ICU	1109 (100)	25 (2)	1084 (98)	
Time period, n (%)					<.001
	After September 1, 2021	6163 (100)	1280 (21)	4883 (79)	
	Until August 31, 2021	7848 (100)	2265 (29)	5583 (71)	
Hospital region, n (%)					<.001
	Attica	5892 (100)	1817 (31)	4075 (69)	
	Thessaloniki	3065 (100)	760 (25)	2305 (75)	
	Rest of Greece	5054 (100)	968 (19)	4086 (81)	
Total number of people intubated (at date of intubation), n (%)					<.001
	0-199	767 (100)	277 (36)	490 (64)	
	200-299	821 (100)	284 (35)	537 (65)	
	300-399	2824 (100)	756 (27)	2068 (73)	
	400-499	1648 (100)	442 (27)	1206 (73)	
	500-599	2055 (100)	459 (22)	1596 (78)	
	600-699	3300 (100)	708 (21)	2592 (79)	
	700-799	1686 (100)	396 (23)	1290 (77)	
	≥800	910 (100)	223 (25)	687 (75)	

[a]The denominator for the percentages in these columns is the value in the Total column.

[b]Mann-Whitney test for continuous variables; Fisher test for categorical variables

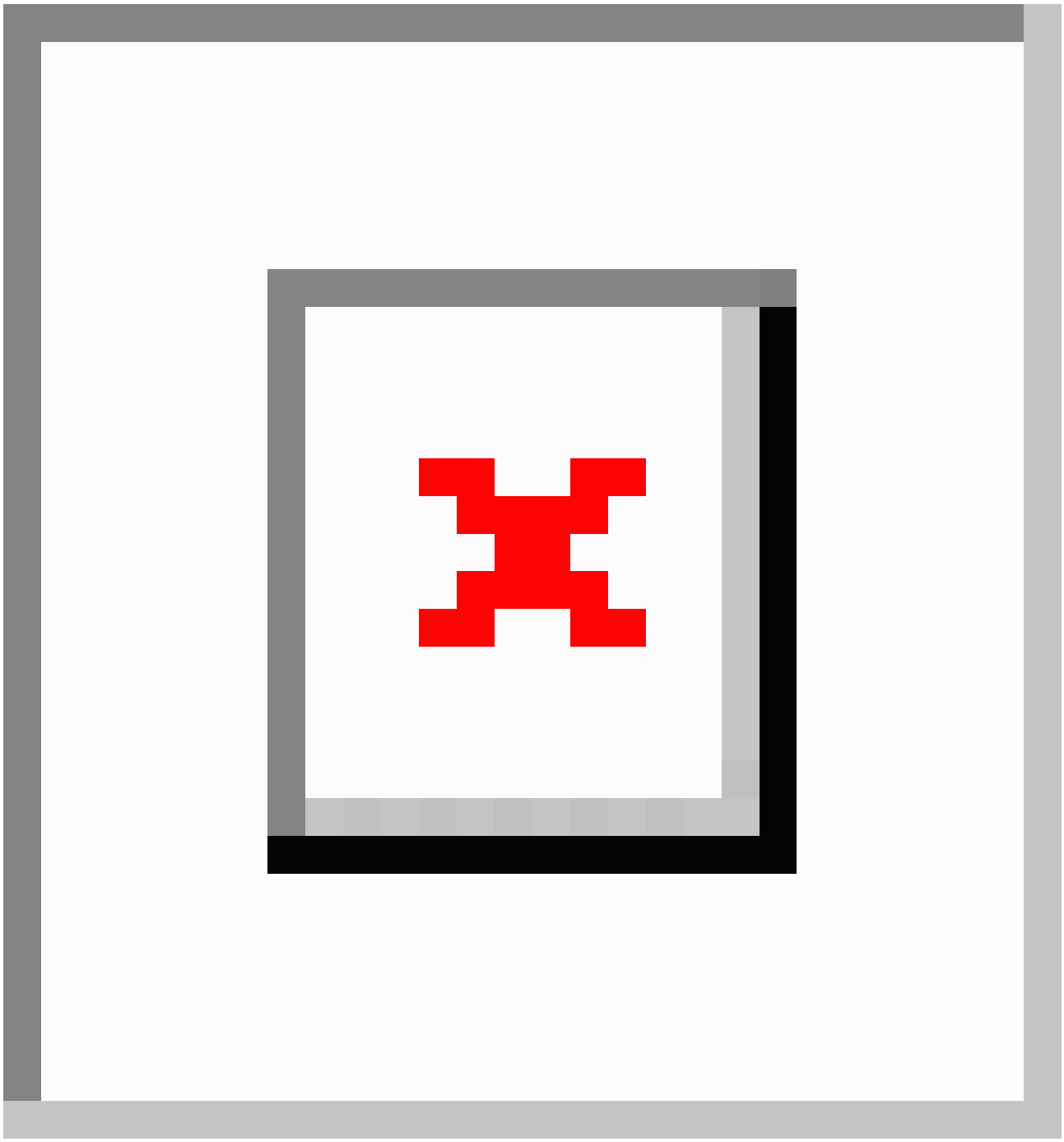
[c]ICU: intensive care unit.

All model results, expressed as adjusted hazard ratios (HRs) and 95% CIs, are shown in Figure 2. There was a significant association between mortality and a total number of intubated patients above 400, with its magnitude increasing progressively: from 1.22 (95% CI 1.09-1.38) for 400-499 patients, up to 1.48 (95% CI 1.31-1.69) for ≥800 patients. Interestingly, for a given patient load there was substantially increased mortality for the period after September 1, 2021, with an HR of 1.21 (95% CI 1.09-1.36). Being hospitalized outside the capital region of Attica was also associated with increased in-hospital mortality, especially for the rest of the country (besides Attica and Thessaloniki) and even more so after September 1, with an HR of 1.64 (95% CI 1.54-1.75). In addition, age was strongly associated with mortality, as was being intubated outside an ICU (HR 2.01, 95% CI 1.89-2.15). Vaccination did not show a statistically significant association with mortality, regardless of the number of doses received (Figure 2).

Given the above associations, of the 10,466 deaths reported, 2176 (95% CI 1297-2986) were attributable to the high load

(≥200) of intubated patients with COVID-19, 564 (95% CI 499-634) to being outside an ICU, 1786 (95% CI 1572-2002) to being hospitalized away from Attica, and 1196 (95% CI 779-1586) to being hospitalized after September 1, 2021, and thus experiencing increased mortality. A combined total of 4677 deaths (95% CI 4021-5284) were attributable to all 4 factors collectively.

Figure 2. Multivariable (adjusted) associations between in-hospital mortality of intubated patients with COVID-19 and age, sex, hospitalization type, hospital region, and patient load in Greece from September 1, 2020, to May 17, 2022. HR: hazard ratio; ICU: intensive care unit.



Discussion

Our analysis confirms the earlier findings of our previously published study, indicating that in-hospital mortality of severely ill patients with COVID-19 is adversely affected by high patient load [3]; indeed, the associations between patient load and mortality were nearly identical to our previous analysis but now with significantly higher precision (narrower 95% CIs) as a result of a much larger sample (14,011 patients compared to 6282 previously). This shows that patient outcomes are affected not just when health care capacity is stressed to depletion, but also at lower levels, despite the availability of care not nominally being restricted. This represents a major preventable factor to limit avoidable deaths from COVID-19, and points toward more extensive investment in preparedness and resilience in health care. We also reconfirmed substantial regional disparities, with in-hospital mortality being lower in Thessaloniki and even lower in Attica compared to the rest of the country, highlighting the

chronically uneven regional distribution of health care resources in Greece [6].

A major new finding is the 21% higher mortality observed in the period starting September 2021, which increased even further for patients hospitalized in the rest of the country (an additional +64%, compared to +36% for the previous period). This suggests that conditions in health care services over the past year may have further deteriorated, especially in rural areas (outside Attica and Thessaloniki). It must be noted that from September 2021 the government suspended those health care workers that remained unvaccinated against COVID-19 (although many of them eventually received vaccinations and returned to duty or were replaced with newly hired personnel). Our data cannot prove a causal association between this disciplinary action and the increased mortality, but the temporal coincidence is still worrisome and merits further exploration into the precise causes of this worsening health care system performance.

Our analysis reconfirms that being intubated outside an ICU is associated with twice higher mortality than inside an ICU, although again this must be interpreted with caution; patient selection (prioritizing those with a higher chance of survival for ICU admission) may account for a large part of this association. On the other hand, we found no statistically significant association between mortality and the number of COVID-19 vaccine doses received; this suggests that despite vaccination being enormously effective in preventing severe COVID-19 and death [7], if a patient has already been intubated as a result of severe COVID-19, it is quality of care and not vaccination that can prevent further deterioration and death.

In fact, the observed nonsignificant trend toward higher mortality among those having received 2 or 3 vaccine doses may be the result of some uncontrolled confounding due to comorbidities, as older people and those with comorbidities tend to be more frequently vaccinated. Indeed, the lack of information on comorbidities and baseline health status is the single and most significant limitation of our analysis; nevertheless, these factors are unlikely to be differentially distributed over time and with different patient loads. Both factors are also eclipsed by age, which is the major determinant of the risk of death and which we carefully adjusted for [8,9]. Therefore, residual confounding by baseline health is unlikely to have affected these results to any substantial extent.

As a result, the observed associations most likely reflect real and avoidable differences in the quality of care for patients with COVID-19 due to increased patient load, regional disparities, and ICU availability, as well as a true deterioration after September 1, 2021. The findings highlight the need for urgent strengthening of health care services in Greece in order to improve their performance and ensure equitable access to high-quality care for all [10].

Acknowledgments

We acknowledge the staff of the National Public Health Organization for their hard work in coordinating the surveillance of COVID-19 in Greece.

Data Availability

Data are available in Multimedia Appendix 1 under a Creative Commons BY-NC license [11], as they were provided to the author by the National Public Health Organization without any restrictions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study data set (CC BY-NC).
[ZIP File, 107 KB - xmed_v5i1e43341_app1.zip]

References

1. Soria A, Galimberti S, Lapadula G, et al. The high volume of patients admitted during the SARS-Cov-2 pandemic has an independent harmful impact on in-hospital mortality from COVID-19. PLoS One 2021 Jan 28;16(1):e0246170. [doi: 10.1371/journal.pone.0246170] [Medline: 33507954]
2. Bravata DM, Perkins AJ, Myers LJ, et al. Association of intensive care unit patient load and demand with mortality rates in US Department of Veterans Affairs hospitals during the COVID-19 pandemic. JAMA Netw Open 2021 Jan 4;4(1):e2034266. [doi: 10.1001/jamanetworkopen.2020.34266] [Medline: 33464319]
3. Lytras T, Tsiodras S. Total patient load, regional disparities and in-hospital mortality of intubated COVID-19 patients in Greece, from September 2020 to May 2021. Scand J Public Health 2022 Aug;50(6):671-675. [doi: 10.1177/14034948211059968] [Medline: 34903101]
4. Carstensen B. Who needs the Cox model anyway? Bendixcarstensen.com. 2019. URL: http://bendixcarstensen.com/WntCma.pdf [accessed 2024-05-21]
5. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. Biometrics 1993 Sep;49(3):865-872. [Medline: 8241375]

6. Economou C, Kaitelidou D, Karanikolos M, Maresso A. Greece: health system review. Health Syst Transit 2017 Sep;19(5):1-192. [Medline: 29972131]
7. Lytras T, Kontopidou F, Lambrou A, Tsiodras S. Comparative effectiveness and durability of COVID-19 vaccination against death and severe disease in an ongoing nationwide mass vaccination campaign. J Med Virol 2022 Oct;94(10):5044-5050. [doi: 10.1002/jmv.27934] [Medline: 35701379]
8. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020 May 22;369:m1966. [doi: 10.1136/bmj.m1966] [Medline: 32444366]
9. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436. [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
10. Paremoer L, Nandi S, Serag H, Baum F. Covid-19 pandemic and the social determinants of health. BMJ 2021 Jan 28;372:129. [doi: 10.1136/bmj.n129] [Medline: 33509801]
11. Attribution-Noncommercial 4.0 International (CC BY-NC 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by-nc/4.0/ [accessed 2024-05-03]

Abbreviations

HR: hazard ratio
ICU: intensive care unit
NPHO: National Public Health Organization

Edited by G Eysenbach, T Leung; submitted 09.10.22; peer-reviewed by A Vighio, Anonymous, M Coccia, N Mathews; revised version received 24.01.24; accepted 26.03.24; published 10.06.24.

Please cite as:
Lytras T
Health Care System Overstretch and In-Hospital Mortality of Intubated Patients With COVID-19 in Greece From September 2020 to April 2022: Updated Retrospective Cohort Study
JMIRx Med 2024;5:e43341
URL: https://xmed.jmir.org/2024/1/e43341
doi:10.2196/43341

Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020

Thokozani Mzumara[1,2,3], MSc; Marios Kantaris[3,4], PhD; Joseph Afonne[1], OD

1

2

3

4

Corresponding Author:
Thokozani Mzumara, MSc

Related Articles:

Companion article: http://preprints.jmir.org/preprint/44381

Companion article: https://med.jmirx.org/2024/1/e58361

Companion article: https://med.jmirx.org/2024/1/e57935

Companion article: https://med.jmirx.org/2024/1/e57620

Abstract

Background: The use of eye care services varies among different population groups.

Objective: This study aimed to assess self-reported eye care use (ECU) and associated demographic factors among Malawian adults.

Methods: This study used secondary data from the Malawi Fifth Integrated Household Survey 2019-2020, a nationally representative survey. The study included 12,288 households and 27,336 individuals 15 years and older. We entered age, sex, level of education, residency (urban/rural), and chronic disease into a logistic regression model, and used a confusion matrix to predict the model's accuracy. A *P* value <.05 was considered statistically significant.

Results: About 60.6% (95% CI 60.0%-61.2%) of those with eye problems accessed formal care 2 weeks before the survey date. A logistic regression model showed that ECU was positively associated with education compared to none (odds ratio [OR] 6.6, 95% CI 5.927-7.366; *P*<.001), males compared to females (OR 1.2, 95% CI 1.104-1.290; *P*<.001), and urban residence compared to rural (OR 1.2, 95% CI 1.118-1.375; *P*<.001). ECU was negatively associated with age (OR 7, 95% CI 6.782-8.476; *P*<.001) and having chronic diseases (OR 0.6, 95% CI 0.547-0.708; *P*<.001).

Conclusions: Social support, women empowerment, education, and mobile clinics are key strategic areas that would increase access to eye care in Malawi. Further studies can investigate ECU among the pediatric population.

(JMIRx Med 2024;5:e44381) doi:10.2196/44381

KEYWORDS

access to health; health service utilization; eye care use; health-seeking behavior; sociodemographic determinant; visual impairment; social support; women empowerment; education; eye care; pediatric; eye; ophthalmology; visual; ECU; eye service; utilization; Malawi; empowerment; health service use; use

Introduction

Life, health, and sustainable development depend on good ocular health, yet many people cannot afford the eye care they need, which can result in blindness or visual impairment [1,2]. Approximately, 90% of the 596 million individuals with visual impairments globally in 2020 lived in low-income nations [3,4]. Access, eye care use (ECU), and poverty are all directly correlated with the prevalence of visual impairment [2]. According to the literature [5], poverty is a direct cause and

effect of blindness. It affects impoverished communities in certain ways and prevents the poor from accessing eye care services because the associated expenses are too high. Socioeconomic variables including education and ethnicity are to blame for disparities in health-seeking behavior (HSB), access to services, visual impairment, and blindness [5]. Blindness exacerbates poverty, which results in diminished autonomy, position, and authority as well as decreased involvement, social isolation, and stigma. This can extend to other family members, leading to despair and decreased productivity. Beyond a lack of money, poverty also includes a loss of control over one's life; a loss of status, authority, and prestige; fewer possibilities to engage in family- and work-related activities; stigma; and social isolation.

ECU entails using eye activity for treatment, prevention, and health promotion [6-8]. Health care use is influenced by sociocultural variables, norms, autonomy, and decision-making. Additionally, human behavior poses a challenge to cost-effective health care solutions [9]. HSB is influenced by the health belief model, driven by accessibility to care, perceived severity, susceptibility, and treatment benefits [10]. The model distinguishes potential and realized access, with predisposing factors like age, sex, and lifestyle; need factors like ill health; and enabling factors like wealth and proximity to health facilities [11]. Nevertheless, HSB is not homogeneous and relies on cognitive and noncognitive contexts [10]. The factors affecting HSB are multifaceted; hence, wide variations are common, making it a challenge to health equity and universal eye health.

Globally, ECU varies from 18% to 83% [6] and is characterized by inequalities based on sex, comorbidity, age, and education [12]. Furthermore, ECU varies between rural and urban residences [12]. Studies have found that visual problems are higher among those less educated and who live in rural areas [5]. Numerous studies have looked into the ECU pattern and the risk variables that are linked to it [6,13-18]. This article attempts to evaluate the pattern of ECU among adults in Malawi, concentrating on related demographic parameters, given that ECU varies with context. The study also looks at the decisions made by individuals who have vision issues, including their reasons for not seeking eye care.

Malawi is among the world's poorest nations [19]. According to a recent study, Malawi must first eradicate inequality before poverty can be eliminated [20]. The results of this study can help address the systematic exclusion of disadvantaged groups and develop health systems toward universal coverage of eye health through evidence-based programs and policies to increase the use of eye care services. The attainment of universal eye health is hindered by low uptake of eye care, especially in low-income countries where access to care is restricted and patients delay seeking treatment due to financial restrictions, fear, or neglect [6,10,21].

Methods

Survey

This quantitative cross-sectional correlation study was conducted as a desk-based review using data obtained from the Malawi Fifth Integrated Health Survey (IHS) 2019-2020 [22]. The survey, a living standards measurement study, was implemented between April 2019 to April 2020. The survey used a nationally representative sample and was conducted across districts in Malawi.

Sampling Procedure

The survey used a stratified 2-stage sampling procedure. The sampling frame was based on listing information and cartography from the 2018 Malawi Population and Housing Census [22], which included the three major regions, namely, the north, center, and south stratified into rural and urban areas. The urban strata included Lilongwe, Mzuzu, Blantyre, and Zomba—as such, the survey considered all other districts to be rural. The study sampled across 32 districts in the nation. The frame excluded populations in institutions such as prisons, hospitals, and military barracks. First, enumeration areas based on the 2018 census were selected within each stratum. Next, households were selected from the EA using systematic sampling. Finally, 12,288 households from 780 enumeration areas were selected with a 93% response rate. The survey interviewed all individuals 15 years and older living in the selected household.

Sampling Weights

The sample estimates from the Fifth IHS were multiplied by sampling weights so that the findings of the study could be extrapolated to the whole population 15 years and older [22].

Data

This review extracted age in years, sex (male/female), residence (rural/urban), highest level of education, and sampling weights from the household module of the survey. The age of participants was regrouped into young adults (15-34 years), middle-aged (35-59 years), and older adults (60 years and older). Participants' level of education was recoded into five categories, namely, none (no education or do not know), primary (primary school leaving certificate), secondary (including A level), and tertiary (diploma, first degree, master's degree, or PhD). In addition, the paper extracted information on whether the individual had experienced any symptoms for the previous 2 weeks (yes/no), symptom names, action taken to relieve the symptom, and having a chronic illness.

To determine ECU, we recorded the action taken to relieve the eye symptom as 1 if the individual sought care at a government or private facility including a church/mission facility, village clinic, and pharmacy store. All other options, including self-care and use of stock medicine and other nonorthodox practices, were recoded as 0.

Analysis Strategy

The data were entered into SPSS, version 26 (IBM Corp). We analyzed descriptive statistics including mean and SD, frequency, and proportion. A graphical illustration of the results is presented in tables. Proportional data were analyzed using χ^2, and we considered a P value <.05 as significant. The variables were entered into a binary logistic regression model using the entry method to estimate the probability of ECU occurring. The probability cutoff was set at 0.5%—probabilities

greater than 0.5 were classified as ECU (event occurring); otherwise, they were classified as no ECU (event not occurring). To assess the prediction classification model, the study used the confusion matrix technique and calculated the predictive values accuracy, specificity, and sensitivity of the model/classifier.

Ethical Considerations

The study adhered to the tenets of the Declaration of Helsinki. Explicitly, the IHS survey obtained informed consent from all participants. Nevertheless, the study did not require institutional review because it used deidentified publicly available data. We obtained permission and data sets from the World Bank [23]. The study subjects participated freely and were not compensated in any form.

Results

Prevalence of Eye Symptoms 2 Weeks Preceding the Survey

Of the 27,336 participants recruited, 8014 (29.3%) were males. The mean age of participants was 47.68 (SD 24.39) years. The proportion of participants with symptoms of eye problems 2 weeks before the study was highest among older adults (60 years and older; n=11,155, 40.8%). According to the region, eye symptoms were higher in the south (n=12,378, 45.3%) and lowest in the north (n=3084, 11.3%). The majority (n=23,202, 84.9%) were from rural areas. The majority of participants with eye symptoms were married (n=14,140, 51.7%; *P*<.001; Table 1).

Table . Characteristics of study participants with eye symptoms.

Characteristics		Participants with eye symptoms reported within the past 2 wk (N=27,336), n (%)
Sex		
	Female	19,323 (70.7)
	Male	8014 (29.3)
Age group		
	Young adults	11,118 (40.7)
	Middle-aged	5063 (18.5)
	Older adults	11,155 (40.8)
Region		
	North	3084 (11.3)
	Central	11,874 (43.4)
	South	12,378 (45.3)
Chronic diseases		
	Yes	2508 (9.2)
	No	24,829 (90.8)
Residence		
	Urban	4135 (15.1)
	Rural	23,301 (84.9)
Marital status		
	Married	14,140 (51.7)
	Separated/divorced	2409 (8.8)
	Widower/widowed	5948 (21.8)
	Not married	4841 (17.7)
Education		
	None	21,032 (76.9)
	Primary	1475 (5.4)
	Secondary	4829 (17.7)

Factors Associated With ECU

The χ^2 test was run as part of the bivariate analysis to assess the association between ECU and various demographic factors. The test showed that a higher proportion of ECU was associated with males than females such that 68.2% (5467/8014) of males sought eye care services compared to 57.4% (11,097/19,322) of females. The percentage of participants who sought care

increased with a higher level of education, as depicted by 53.6% (11,283/21,033) of participants without education who sought care, while 79.1% (1166/1475) of participants with primary education and over 85% (4115/4828) of participants with secondary school education (P<.001) used eye care services. Pearson χ^2 showed that the difference was statistically significant (P<.001). According to place of residence, 66.1% (2733/4134) of urban residents sought care compared to 59.6% (13,831/23,202) of rural residents. The difference was statistically different (P<.001). According to the presence of chronic illness, 77.7% (1949/2508) of participants who had a chronic condition sought care while 88.2% (14,616/24,829) of those participants with no chronic illness sought care (P<.001). Based on the region of residence, 42.7% (1318/3084), 50.4% (5992/11,875), and 74.8% (9254/12,378) of the participants sought care from the north, central, and southern regions, respectively (P<.001). Considering marital status, 13.8% (2286/4840) of participants who were not married sought care, while 61.8% (10,239/14,140) of married people sought care. Among those divorced, only 40.1% (968/2409) of participants sought care and 18.5% (3072/5948) of those who sought care were widowers/widows (P<.001). Regarding age, 54.9% (6110/11,118) of young adults sought care, while 90.2% (4569/5063) of middle-aged adults and 52.7% (5885/11,155) of older adults used eye care services (P<.001).

Places Where Participants Sought Help

A total of 16,564 (60.6%, 95% CI 60.0%-61.2%) of the 27,336 participants sought eye care from a medical/health facility. Of the 16,564 participants who sought care from a health facility, 14,173 (85.6%) visited a government facility, and 463 (2.7%) obtained drugs from their local pharmacy. Among those who did not seek care, 2950 of 6343 (46.5%) attributed it to lack of funds, while 3393 of 6343 (53.5%) did not think it was a serious illness (Table 2).

Table . Distribution of actions taken by participants with eye symptoms 2 weeks before the survey.

			Participants (N=27,366), n (%)
Did not seek health care[a]			6343 (23.1)
	Did nothing, not a serious illness		3393 (53.4)
	Did nothing, no money		2950 (46.8)
Sought health care			
	C are at health facility[b]		16,565 (60.5)
		Government	14,173 (85.5)
		Private	1316 (7.9)
		Church/mission	613 (3.7)
		Pharmacy store	463 (2.7)
	Care not at health facility[a]		3788 (13.8)
		Personal remedies	2057 (54.3)
		Grocery store	1123 (29.6)
		Traditional healer	608 (16.2)

[a]These actions do not constitute eye care use.

[b]These actions constitute eye care use.

Factors Affecting ECU

A logistic regression was performed to ascertain the effects of residence, region, education qualification, chronic illness, sex, and age on the likelihood that participants reported ECU. The logistic regression model was statistically significant (χ^2_8=27.4; P<.001). The model explained 34.0% (Nagelkerke R^2) of the variance in ECU. Males were more likely to report ECU than females (odds ratio [OR] 1.2, 95% CI 1.104-1.290). Having a chronic condition was associated with a reduction in the likelihood of ECU (OR 0.6, 95% CI 0.547-0.708). Residents in the urban area were 1.2 times more likely to exhibit ECU than residents in rural areas (OR 1.2, 95% CI 1.118-1.375). Those with a higher education qualification were 6 times more likely to seek eye care at a medical facility than those without a formal education (OR 6.6, 95% CI 5.927-7.366). Middle-aged participants were 7 times more likely to use eye care services than young adults (OR 7, 95% CI 6.782-8.476), while older adults were less likely to use eye care than middle-aged adults.

Discussion

The prevalence of ECU in our study was similar to a previous report [11]. However, other authors report comparatively lower rates of ECU [6,15]. On the contrary, others reported larger ECU rates [16,24]. There is a discrepancy in the rate of ECU as it varies widely ranging from 18% to 82% due to study settings, sample size, and study population [6]. In part, the variation could be due to different operational definitions of ECU. For instance, we defined ECU as seeking care for ocular problems at a medical facility 2 weeks before the study, whereas the American Optometric Association defines ECU as the use of eye care services in the preceding 3 years [25]. The high rate

of ECU in our study could be attributed to the affordability of eye care services in the country [26].

The majority of participants who sought care in our study visited government hospitals. On the contrary, Kyaw et al [27] found that the most frequently visited place is private hospitals, and Morka et al [6] reported that the most preferred place of choice is an eye center [28]. The results vary depending on the study setting. The results of our study are not surprising considering that the government is the chief actor in health service delivery in the country [26].

Regarding the reasons for not seeking care, we found that the majority of participants cited a lack of funds. Cost is a common barrier in many low-income nations [29,30]. Although eye care is provided free of charge in Malawi, individuals still incur pocket expenditures in the form of transport costs [31]. Assessing barriers to ECU is beyond the scope of this paper; nevertheless, our finding endorses the need to scale up outreach programs to mitigate costs incurred when reaching a health facility.

Another larger group of people did not seek care for their eye problems due to "negligence" similar to the previous report [30]. This reflects poor-functioning eye health awareness programs and could be a contributing factor to the late presentation of ocular conditions to the hospital. Therefore, the findings of this study suggest that community sensitization and eye health education programs emphasize the gravity of eye problems and the significance of early presentation to the hospital.

A previous report noted that the prevalence of self-medication was 40% [32]; on the contrary, this study found that a decreased number of participants resorted to self-medication. The difference could be due to different sample sizes. Our data used a nationwide sample, unlike the previous study that recruited in only two districts. Regardless, self-medication is a custom in Africa where the majority of eye problems do not go beyond unorthodox alternatives [32]. Our study underscores the significance of incorporating nonorthodox approaches into the national eye health system.

Concerning age, ECU was highest among young adults but dropped among older adults despite a high prevalence of eye symptoms. The trend is similar to Zhao and colleagues [33]; however, it is in disagreement with others [6,34]. In general, major causes of eye diseases are age related; thus, demand for eye services increases with age [6]. The results of our study could be attributed to a lack of social support. Research suggests decreased ECU among older adults due to the lack of an escort to the health facility [6]. This highlights an inefficient eye care delivery system with top-heavy needs and supply. Hence, we recommend gearing services toward older adults, such as the aforementioned vision-screening programs oriented toward senior citizens.

Arguably, one's cultural values have an impact on access to health. Our study has shown that male subjects are associated with ECU more than females, similar to previous authors [35,36]. In contrast, others report females' predisposition to ECU [33,34]. Nonetheless, Akuwoah and colleagues [29] found no statistically significant difference between genders. The variation can be explained by different cultural backgrounds in different populations. The high ECU rate among males in our study can be explained by the Malawian cultural scene. Researchers demonstrated that Malawian women's underuse of necessary health care was a factor in decision-making and attitude [8]. We advocate for women's empowerment and mitigation of gender stereotypes to improve equal access to eye care services in the country. The literature suggests that education is instrumental in curbing such gender disparities [8].

Generally, education is a key affluence factor in health as it directly relates to awareness [37]. Our investigation has shown that the rate of ECU is higher among those with more education similar to previous studies [34,37]. This provides an attractive target for strategies to improve ECU through measures to keep citizens in school, promote equal education opportunities, and design eye health messages targeting citizens who are illiterate.

Surprisingly, this review has shown that having a chronic disease is negatively associated with ECU. However, other researchers [16,34] found that the absence of diabetes was associated with low ECU. Etiologically, most ocular conditions are caused by chronic diseases [11]. Given that the study was cross-sectional and that participants had just 2 weeks to report eye issues, it is crucial to note that chronic illnesses could be more common than reported in this study. Moreover, we attributed the results of our study to a lack of knowledge and awareness. A recent study found that understanding health and disease is a key determinant of seeking health care among persons with chronic conditions in Malawi [38]. When health education messages demonstrate that there is a potential health risk and convince people that certain behaviors can prevent such risks, the likelihood of change is increased substantially [10]. An imperative but unanswered question is how individual chronic diseases affect ECU. Regardless, our findings echo the significance of awareness campaigns among people with chronic diseases.

Our investigation revealed that people from urban areas use eye care services more than their rural counterparts comparable to previous studies [24]. The rural-urban disparity can be explained by less available ophthalmic human resources in rural Malawi. For instance, 11 of 12 ophthalmologists in the country are based in urban areas [39]. The results of the study suggest inequity in deployment policies; hence, task-shifting approaches would ensure coverage in hard-to-reach areas.

Concerning region of origin, ECU in Malawi varies, with the south registering the highest and the north having the lowest levels of ECU. We cannot explain the regional variation in our study.

A particular strength of this study is the use of a countrywide population-based data set that provides ample sample size and statistical power to reconnoiter the association between ECU and its associated factors. Nevertheless, our study is not without drawbacks. First, the study is based on subjective responses that are prone to recall bias by the study participants. Another important caveat of our study is that we did not include the association between economic status and ECU including wealth and out-of-pocket expenditure. In addition, we did not evaluate

realized access, patient satisfaction, and quality of care due to a lack of objective measures.

In summary, we report a novel study on the rate of ECU and its associated demographic factors among a population of Malawian adults using data from a national survey. The results demonstrate that in Malawi access to eye care services is entrenched in social disadvantages. ECU is low among older adults due to a lack of social support. Unequal distribution of resources between urban and rural areas could cause disparities in ECUs. There were regional variations in the use of eye services. Measures and interventions to improve ECU should target strategies to increase education opportunities for all, women empowerment, and outreach programs for rural residents and older adults, including task-shifting programs.

Data Availability

Publicly available data sets [23] were analyzed in this study.

Conflicts of Interest

None declared.

References

1. Burton MJ, Ramke J, Marques AP, et al. The Lancet Global Health Commission on global eye health: vision beyond 2020. Lancet Glob Health 2021 Apr;9(4):e489-e551. [doi: 10.1016/S2214-109X(20)30488-5] [Medline: 33607016]
2. Rono HK, Macleod D, Bastawrous A, Wanjala E, Gichangi M, Burton MJ. Utilization of secondary eye care services in Western Kenya. Int J Environ Res Public Health 2019 Sep 12;16(18):3371. [doi: 10.3390/ijerph16183371] [Medline: 31547252]
3. Kalua K, Lindfield R, Mtupanyama M, Mtumodzi D, Msiska V. Findings from a rapid assessment of avoidable blindness (RAAB). PLoS One 2011 Apr 25;6(4):e19226. [doi: 10.1371/journal.pone.0019226] [Medline: 21547074]
4. Flaxman AD, Wittenborn JS, Robalik T, et al. Prevalence of visual acuity loss or blindness in the US: a Bayesian meta-analysis. JAMA Ophthalmol 2021 Jul 1;139(7):717-723. [doi: 10.1001/jamaophthalmol.2021.0527] [Medline: 33983373]
5. Tafida A, Kyari F, Abdull MM, et al. Poverty and blindness in Nigeria: results from the national survey of blindness and visual impairment. Ophthalmic Epidemiol 2015;22(5):333-341. [doi: 10.3109/09286586.2015.1077259] [Medline: 26395660]
6. Morka ED, Yibekal BT, Tegegne MM. Eye care service utilization and associated factors among older adults in Hawassa City, South Ethiopia. PLoS One 2020 Apr 16;15(4):e0231616. [doi: 10.1371/journal.pone.0231616] [Medline: 32298344]
7. Kegne Assaye A, Temeselew Tegegn M, Tilahun Belete G. Eye care utilization among older subjects with visual impairment in Northwest Ethiopia. J Ophthalmic Vis Res 2023 Jul 28;18(3):306-317. [doi: 10.18502/jovr.v18i3.13779] [Medline: 37600912]
8. Gellman MD. Behavioral medicine. In: Gellman MD, editor. Encyclopedia of Behavioral Medicine: Springer International Publishing; 2020:223-226. [doi: 10.1007/978-3-030-39903-0]
9. Azad AD, Charles AG, Ding Q, Trickey AW, Wren SM. The gender gap and healthcare: associations between gender roles and factors affecting healthcare access in central Malawi, June-August 2017. Arch Public Health 2020 Nov 17;78(1):119. [doi: 10.1186/s13690-020-00497-w] [Medline: 33292511]
10. Oberoi S, Chaudhary N, Patnaik S, Singh A. Understanding health seeking behavior. J Family Med Prim Care Apr-Jun 2016;5(2):463-464. [doi: 10.4103/2249-4863.192376] [Medline: 27843863]
11. Akuffo KO, Sewpaul R, Dukhi N, et al. Eye care utilization pattern in South Africa: results from SANHANES-1. BMC Health Serv Res 2020 Aug 17;20(1):756. [doi: 10.1186/s12913-020-05621-8] [Medline: 32807155]
12. Etim BA, Ibanga AA, Nkanga DG, Agweye CT, Utam UA, Udofia OO. EHR health seeking behavior of patients attending eye clinic in Southern Nigeria. Niger J Clin Pract 2019 Jul;22(7):988-996. [doi: 10.4103/njcp.njcp_373_18] [Medline: 31293266]
13. Jin YP, Trope GE. Eye care utilization in Canada: disparity in the publicly funded health care system. Can J Ophthalmol 2011 Apr;46(2):133-138. [doi: 10.3129/i10-120] [Medline: 21708079]
14. Taccheri C, Jordan J, Tran D, et al. The impact of social determinants of health on eye care utilization in a national sample of people with diabetes. Ophthalmology 2023 Oct;130(10):1037-1045. [doi: 10.1016/j.ophtha.2023.06.007] [Medline: 37329902]
15. Andoh JE, Ezekwesili AC, Nwanyanwu K, Elam A. Disparities in eye care access and utilization: a narrative review. Annu Rev Vis Sci 2023 Sep 15;9:15-37. [doi: 10.1146/annurev-vision-112122-020934] [Medline: 37254050]
16. Ezinne NE, Ekemiri KK, Chukwuma I, et al. Utilization of eye care services in an underserved community in Enugu State, Nigeria. Niger J Clin Pract 2023 Jan;26(1):81-89. [doi: 10.4103/njcp.njcp_398_22] [Medline: 36751828]
17. Lin JC, Ghauri SY, Scott IU, Greenberg PB. Eye health care utilization among native Hawaiian and Pacific Islander adults in the United States. Ophthalmic Epidemiol 2022 Feb 2:1-5. [doi: 10.1080/09286586.2022.2036765] [Medline: 35109764]

18. Pallavi DR, Mohit Bohra D, Kauser H, Talawat M. An empirical survey on eye care services in Bengaluru. REST J Data Analytics Artif Intelligence 2023 Jun;2(2):1-16 [FREE Full text] [doi: 10.46632/jdaai/2/2/1]
19. Masefield SC, Msosa A, Grugel J. Challenges to effective governance in a low income healthcare system: a qualitative study of stakeholder perceptions in Malawi. BMC Health Serv Res 2020 Dec 14;20(1):1142. [doi: 10.1186/s12913-020-06002-x] [Medline: 33317520]
20. Mussa R, Masanjala WH. A dangerous divide: the state of inequality in Malawi. Oxfam Digital Repository. 2015 Nov 26. URL: https://oxfamilibrary.openrepository.com/handle/10546/582678 [accessed 2024-03-20]
21. Ebeigbe JA, Ovenseri-Ogbomo GO. Barriers to utilization of eye care services in rural communities in Edo state, Nigeria. Bo Med J 2014;11(2):98-104 [FREE Full text]
22. Ng'ambi W, Mangal T, Phillips A, et al. A cross-sectional study on factors associated with health seeking behaviour of Malawians aged 15+ years in 2016. Malawi Med J 2020 Dec;32(4):205-212. [doi: 10.4314/mmj.v32i4.5] [Medline: 34457205]
23. Integrated Household Panel Survey 2010-2013-2016 (Long-Term Panel, 102 EAs). The World Bank: Microdata Library. URL: https://microdata.worldbank.org/index.php/catalog/2939 [accessed 2024-03-21]
24. Foreman J, Xie J, Keel S, Taylor HR, Dirani M. Utilization of eye health-care services in Australia: the national eye health survey. Clin Exp Ophthalmol 2018 Apr;46(3):213-221 [FREE Full text] [doi: 10.1111/ceo.13035] [Medline: 28793183]
25. Ilechie AA, Otchere H, Darko-Takyi C, Halladay AC. Access to and utilization of eye care services in Ghana. Int J Health Res 2013 Sep;6(3):7-15 [FREE Full text]
26. Mugwang'a Z, Bechange S, Masika M. Eye care service assessment: Malawi. Sightsavers. 2017. URL: https://research.sightsavers.org/wp-content/uploads/2018/05/Eye-care-service-assessment-Malawi-2017.pdf [accessed 2024-04-01]
27. Kyaw KK, Nanthamonkolchai S, Munsawaengsub C. Eye care seeking behavior on prevention of blindness among elderly in urban area of Magway Township, Myanmar. Int J Ophthalmol Eye Sci 2018;6(1):351-357. [doi: 10.19070/2332-290X-1800071]
28. National Statistical Office. Fifth Integrated Household Survey 2019-2020: Malawi, 2019-2020. The World Bank: Mircodata Library. 2020 Dec 1. URL: https://microdata.worldbank.org/index.php/catalog/3818/get-microdata [accessed 2021-12-03]
29. Akowuah PK, Merepa SS, Abazele AS. Barriers to utilization of eye care services in the upper east region, Ghana. Adv Ophthalmol Vis Syst 2017 Nov 16;7(6). [doi: 10.15406/aovs.2017.07.00240]
30. Arinze OC, Eze BI, Ude NN, Onwubiko SN, Ezisi CN, Chuka-Okosa CM. Determinants of eye care utilization in rural southeastern Nigeria. J Community Health 2015 Oct;40(5):881-890. [doi: 10.1007/s10900-015-0008-2] [Medline: 25787225]
31. Ng'ambi W, Mangal T, Phillips A, et al. Factors associated with healthcare seeking behaviour for children in Malawi: 2016. Trop Med Int Health 2020 Dec;25(12):1486-1495. [doi: 10.1111/tmi.13499] [Medline: 32981174]
32. Bisika T, Courtright P, Geneau R, Kasote A, Chimombo L, Chirambo M. Self treatment of eye diseases in Malawi. Afr J Tradit Complement Altern Med 2008 Oct 25;6(1):23-29. [doi: 10.4314/ajtcam.v6i1.57070] [Medline: 20162038]
33. Zhao M, Gillani AH, Amirul Islam FM, et al. Factors associated with knowledge, attitude and practices of common eye diseases in general population: a multicenter cross-sectional study from Pakistan. Int J Environ Res Public Health 2019 May 5;16(9):1568. [doi: 10.3390/ijerph16091568] [Medline: 31060273]
34. Jiang X, Varma R, Torres M, Hsu C, McKean-Cowdin R, Chinese American Eye Study Group. Self-reported use of eye care among adult Chinese Americans: the Chinese American eye study. Am J Ophthalmol 2017 Apr;176:183-193. [doi: 10.1016/j.ajo.2017.01.018] [Medline: 28161048]
35. Ocansey S, Kumi-Kyereme A, Awusabo-Asare K, Ilechie A, Boadi-Kusi S, Abraham C. Utilization of eye care services among Ghanaian elderly population: evidence from a peri-urban community. Ophthal Res Int J 2013 Sep 27;1(2):89-101. [doi: 10.9734/OR/2013/5543]
36. Olusanya BA, Ashaye AO, Owoaje ET, Baiyeroju AM, Ajayi BG. Determinants of utilization of eye care services in a rural adult population of a developing country. Middle East Afr J Ophthalmol Jan-Mar 2016;23(1):96-103. [doi: 10.4103/0974-9233.164621] [Medline: 26957847]
37. Raznahan M, Emamian MH, Hashemi H, Zeraati H, Fotouhi A. Assessment of horizontal inequity in eye care utilization in the Iranian middle-aged population. J Ophthalmic Vis Res Jul-Sep 2018;13(3):284-292. [doi: 10.4103/jovr.jovr_221_17] [Medline: 30090185]
38. Saleh S, Bongololo G, Banda H, et al. Health seeking for chronic lung disease in central Malawi: adapting existing models using insights from a qualitative study. PLoS One 2018 Dec 17;13(12):e0208188. [doi: 10.1371/journal.pone.0208188] [Medline: 30557307]
39. Kalua K. How to create a balanced eye team: an example from Malawi. Community Eye Health 2018;31(102):46. [Medline: 30220802]

Abbreviations

ECU: eye care use
HSB: health-seeking behavior
IHS: Integrated Health Survey

OR: odds ratio

Edited by E Meinert; submitted 17.11.22; peer-reviewed by Anonymous, Anonymous; revised version received 07.01.24; accepted 19.02.24; published 09.04.24.

Please cite as:
Mzumara T, Kantaris M, Afonne J
Eye Care Service Use and Associated Health-Seeking Behaviors Among Malawian Adults: Secondary Analysis of the Malawi Fifth Integrated Household Survey 2019-2020
JMIRx Med 2024;5:e44381
URL: https://xmed.jmir.org/2024/1/e44381
doi:10.2196/44381

Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study

Zaeem Ul Haq[1,*], PhD; Ayesha Naeem[2,*], MA; Durayya Zaeem[3], BSc; Mohina Sohail[4], MD; Noor ul Ain Pervaiz[4], MA

1

2

3

4

*these authors contributed equally

Corresponding Author:
Zaeem Ul Haq, PhD

Related Articles:

Companion article: https://preprints.jmir.org/preprint/48213

Companion article: https://med.jmirx.org/2024/1/e60430

Companion article: https://med.jmirx.org/2024/1/e60429

Companion article: https://med.jmirx.org/2024/1/e60266

Abstract

Background: Primary health care (PHC) is the backbone of universal health coverage, with community health workers (CHWs) being one of its critical pillars in lower-middle-income countries. Most CHW functions require them to be an efficient communicator, but their program development has been deficient in this area. Can IT provide some solutions? Moreover, can some IT-based CHW-delivered innovations help mothers and children in areas not covered by PHC services? We explored these questions during the development and feasibility testing of a digital application designed to improve the communication capacity of CHWs in two underserved areas of Islamabad.

Objective: This study aims to explore the perceptions, practices, and related gaps about mother and child health, and child development in an underserved area; develop and deploy a behavior change communication program to address the gaps; and assess the feasibility of the program.

Methods: We carried out a mixed methods study with three steps. First, we conducted 13 in-depth interviews and two focus group discussions with stakeholders to explore the issues faced by mothers living in these underserved areas. To address these barriers, we developed Sehat Ghar, a video-based health education application to demonstrate practices mothers and families needed to adopt. Second, we trained 10 volunteer CHWs from the same community to deliver health education using the application and assessed their pre-post knowledge and skills. Third, these CHWs visited pregnant and lactating mothers in the community with random observation of their work by a supporting supervisor.

Results: Initial exploration revealed a need for health-related knowledge among mothers and suboptimal utilization of public health care. Sehat Ghar used behavior change techniques, including knowledge transfer, enhancing mothers' self-efficacy, and improving family involvement in mother and child care. Volunteer CHWs were identified from the community, who after the training, showed a significant improvement in mean knowledge score (before: mean 8.00, SD 1.49; after: mean 11.40, SD 1.43; $P<.001$) about health. During supportive supervision, these CHWs were rated as excellent in their interaction with mothers and excellent or very good in using the application. The CHW and her community reported their satisfaction with the application and wanted its delivery regularly.

Conclusions: Sehat Ghar is a simple, easy-to-use digital application for CHWs and is acceptable to the community. Mothers appreciate the content and presentation and are ready to incorporate its messages into their daily practices. The real-world effectiveness of the innovation tested on 250 mother-infant pairs will be important for its proof of effectiveness. With its usefulness

and adaptability, and the rapidly spreading use of mobile phones and internet technology, this cost-effective innovation can help in delivering health communications at a large scale in a minimum amount of time.

(JMIRx Med 2024;5:e48213) doi:10.2196/48213

KEYWORDS

primary health care; mother and child health; community health worker; slums; digital applications; health communication.

Introduction

Primary health care (PHC) is the backbone of universal health coverage, which in turn, is critical in the agenda of Sustainable Development Goals [1]. COVID-19 also highlighted the significance of PHC—countries with strong PHC outbreak response systems were more successful in fighting the pandemic [2]. PHC has several essential elements, community health workers (CHWs) being one of them, especially in lower-middle-income countries. Acting as an extension of the health facility, these workers have linkages with families and are responsible for health education, essential medication, and referral for the cases that need to be assessed at the health facility [3].

The performance of these CHWs is crucial for the effectiveness of several dimensions of a health system [4,5]. For example, efficient delivery of health education improves community health behaviors that reduce disease burden and overall health expenditures. Early recognition of symptoms and provision of essential medication by CHWs at the household prevents a family's need to visit their health facility, saving money and avoiding crowding at the facility [6]. Timely identification and referral of complicated cases save lives and improve health [7]. All these functions require the CHW to be an effective communicator. Evaluations of CHW programs, however, usually ignore this primary function [8,9].

Researchers have focused on using digital and mobile phone technology to improve CHW programs. Digital and mobile health interventions have been tested, including those for data collection, reminder SMS text messaging, online training, and delivery of video content on mobile phones [10]. The information gleaned from these studies is mixed. For example, a need for more theoretical models has been prominent in the discourse about interventions for behavior change [11,12]. Moreover, incomplete information from the formative and process evaluation of these studies makes their upscale a challenge. Some have also questioned the usefulness of these technology-based interventions because of the costs, dependence on internet availability, and the training requirements [13].

We explore the answer to these questions by developing and pilot-testing a CHW-delivered digital intervention for populations not covered by PHC in Pakistan. According to the most recent population census in 2023, about 39% of the population in Pakistan lives in urban areas, 56% of which is in underserved areas and squatter settlements [14,15]. These informal settlements usually fall outside the mainstream health and governance system, which is already stretched and unable to provide for even the well-settled populations [16,17]. Women and children face further health challenges because of gender intersectionality [18]. Biomedical interventions to address maternal and child survival and health requirements are available but with partial coverage [19,20].

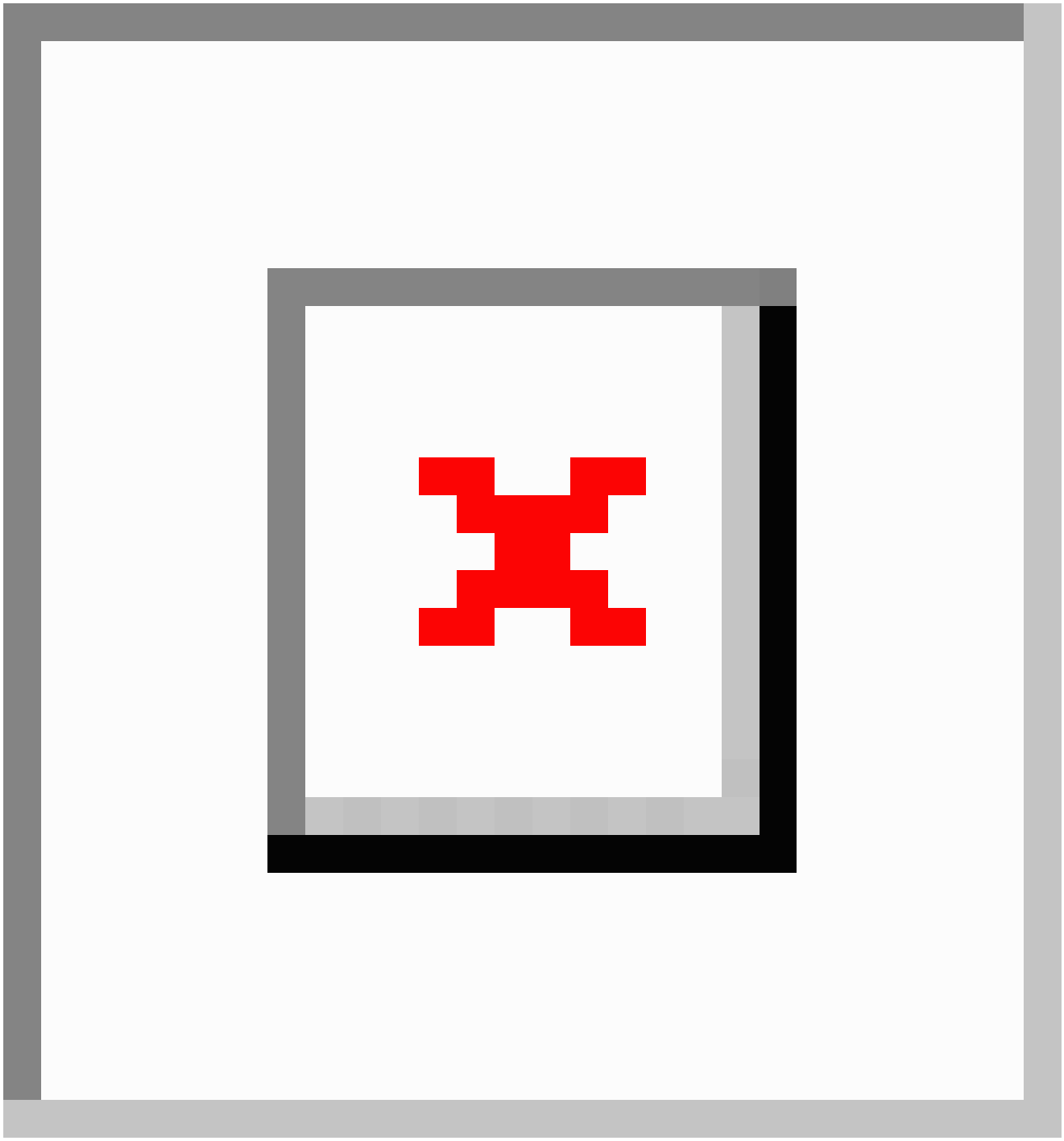
In this paper, we report on the development and feasibility testing of Sehat Ghar (Urdu for "health house"), an Android-based digital application. Volunteer CHWs from the local area are using this application to promote mother and child health among the families living in 2 of the 34 squatter settlements of Islamabad, Pakistan. This formative study aims to pilot-test a tailored health communication intervention that fulfills the knowledge and skill-building needs of a community not covered by the PHC services. The findings will also help other researchers aiming to use digital technology to empower communities and strengthen PHC systems.

Methods

Setting, Design, and Study Objectives

We conducted a mixed methods participatory study (Figure 1) in line with our earlier community-based formative studies [21]. The study involved two underserved areas of Islamabad, Pakistan's capital, with a geographical area of 906 km^2 and a population of over two million [14]. The city has at least 34 underserved areas inhabited by 85,000 people [22]. Many of these underserved areas have poor water, sanitation, and hygiene conditions, and are devoid of public-funded PHC services [23]. We focused our study on France and the Rimsha colonies, two squatter settlements located close to the city's center, with a population of 8000 and 10,000, respectively [24]. We selected these colonies because they fulfilled our criteria of a minimum population of 5000, having a distance of 3 - 5 kilometers from a functioning public hospital where cases could be referred, and health care providers (HCPs) willing to support our study.

Figure 1. Three phases of Sehat Ghar: design, deployment, and implementation. CHW: community health worker.



The residents of both areas mainly belong to the Christian (minority) faith and laborer class, and have low socioeconomic status and poor coverage of PHC services. Our preliminary visits to these areas revealed that there was no significant presence of CHWs from the public sector in these areas. People visited private health providers (both formal and informal) for treatment of minor ailments, and either of the two public hospitals (Polyclinic and Pakistan Institute of Medical Sciences) for a serious condition or hospitalization. Water-borne and other environment-related health problems were common, with mothers and children impacted the most in these areas [17,23].

The following were the specific objectives of this study:

- Explore the perceptions, practices, and related gaps about mother and child health, and child development
- Develop and deploy a behavior change communication program to address the gaps
- Assess the feasibility of the program

Study Participants

We contacted all practicing HCPs in the area—formal and informal—and offered them involvement in our study as a stakeholder. Of the 9 HCPs, 5 (n=2 from Rimsha and n=3 from France colony) agreed. Through them, we identified volunteer women from the community willing to become a CHW. Women were preferred because they could easily visit the households and talk about issues like birth preparedness and complication

readiness [25,26]. In addition to being female, 10 years of schooling, living in the same community, and preferably being married were the criteria for their selection. A total of 15 volunteer women were enrolled. Upon our request, these volunteers conducted household visits and enlisted pregnant or nursing mothers in their area. We invited 10 mothers (5 from each cluster from the list) for in-depth interviews (IDIs) as part of this formative study. Overall, we interviewed 5 HCPs and 8 mothers (2 did not agree to interview), and conducted two focus group discussions (FGDs) with 15 volunteer women (future CHWs for our study) during the inception phase.

Data Collection

Data were collected during all three phases (Figure 1) of the study. Our overarching question was, "In an underserved area, how do pregnant women and children access healthcare, and what are the gaps at the household level that a behavior change intervention can address?" During phase 1, qualitative data were collected to explore the HCP views about the health-seeking practices of their community during the antenatal, natal, and postnatal periods. In the FGDs with future CHWs, we explored the community practices regarding mother and child health, child development, and the gender dimensions prevalent in that culture. Through the IDIs with mothers, we explored the family practices around mother and child nutrition, health-seeking, and child development.

During phase 2, CHWs received training and their knowledge was tested before and after their training, using a yes/no questionnaire. During phase 3, the CHWs visited mothers in their community; their engagement with mothers and families was observed and recorded. A checklist with 21 items (yes/no) was developed for this purpose to record observations on *fidelity to intervention* (6 items), *mastery of using the tablet* (5 items), and *command over the content* (10 items). All CHWs knew that a supervisor would visit them during their household sessions, but the date and time were not revealed to them. Toward the end of a visit, the supervisor also sat with the mother and explored her qualitative feedback using open-ended questions.

For qualitative discussions, two team members with extensive qualitative research experience conducted the interviews and FGDs in Urdu, a language convenient to the community and the study team. The interviews of HCPs and mothers were conducted at the clinic and at home, respectively, while the FGDs were conducted, with consent, in the courtyard of a CHW house. The IDIs took about 30 - 50 minutes, and a typical FGD was 90 minutes. All discussions and interviews were tape-recorded, transcribed, and translated into English.

Data Management and Analysis

For qualitative data, the team transcribed the Urdu discussions from interviews and FGDs and translated these into English. Two team members independently read and coded the initial five transcripts to develop a code list. This initial code list was discussed to see congruence between the two coders and resolve disagreements to reach a final, consensus code sheet. This final code sheet was then used for all interviews and discussions. The analysis identified significant statements across all transcripts using inductive techniques [27,28]. These meaning statements were clustered together to identify themes. We gave weightage to recurring themes; however, we also paid attention to the divergent themes or points that were not shared by many participants but appeared significant [29]. For quantitative analysis, we compared mean scores from the knowledge test of CHWs using a paired 1-tailed *t* test for the hypothesized increase in CHW knowledge after the training. The quantitative data from implementation monitoring were rank-ordered to make supportive supervision decisions during the implementation.

Ethical Considerations

The study team consisted of researchers with medical, public health, and sociology backgrounds; training in mixed methods research; and experience in formative and process evaluations. Most of the team members were female, as the nature of work required close collaboration with mothers and other female members of their households. No incentives or payments were offered to the participants of this research at any stage. To ensure confidentiality and privacy, all data were deidentified for presentation. The Institutional Review Board of the Health Services Academy, Islamabad, Pakistan, approved the study (7 - 82/IERC-HSA/2018 - 06). Grand Challenges Canada funded this study through a seed grant (ST-POC-1808 - 17445), and this formative phase was completed from April to August 2018.

Intervention

In their discussions, the study participants identified knowledge gaps about mother and child health and child development. The discussions also revealed a lack of self-efficacy among mothers to organize their actions around better health outcomes. A digital platform was developed to address the knowledge gaps and provide a platform for the CHW to help the mother and her family through discussions. Applying human-centered design principles [30,31], a multidisciplinary team developed this intervention through several stages. To begin with, the team summarized the findings from stakeholder consultations in the form of a matrix (Table 1) to highlight the community's existing knowledge and skills and relevant gaps. Priority behaviors were identified with the help of HCPs and CHWs, and health education content was finalized to address the gaps.

Table . Gaps in community practices and the intervention content to address the gaps.

Themes	Findings	Intervention element
Access to primary health care	• No coverage of PHC[a] at the household level • Mothers unaware of healthful practices at household level • Families trapped in poverty	• Volunteer CHW[b] introduced as substitute • CHW received basic training to deliver the intervention • Emphasis on behaviors that can prevent disease, promote health, and save costs
Mother's health and nutrition	• Lack of knowledge about antenatal and postnatal care, and regular checkups • Lack of knowledge about maternal nutrition, infrequent use of iron tablets • Mental stress during pregnancy	• Sessions 1 - 3 on antenatal and postnatal checkups, danger signs, and steps to be taken if danger signs appear • Emphasis on maternal nutrition, especially iron/folate • Family support appreciated and further emphasized especially to husband

Newborn health	• Families opt for formal health facility, yet some also prefer TBA[c] for childbirth
	• Newborns receive checkup only if delivered at health facility
	• Prelacteals are common; breastfeeding delayed for 1 - 3 days
	• Warm *desi ghee* (butter) applied to umbilical cord
		• Specific information about SBA[d] and their importance
		• Session 3 on newborn care:
		  • Significance of early checkup
		  • Identifying and taking action for danger signs in a newborn
		  • Early initiation of breastfeeding, delayed bathing for 24 hours
		  • Using chlorhexidine for cord care emphasized

Child health and nutrition	• Children taken to hospital only for immunization or in case of emergency • Lack of knowledge about duration and significance of exclusive breastfeeding • Perception that breastmilk is insufficient • Water and semisolids frequently used between 2 - 5 months • Lack of knowledge about child nutrition	• Sessions 3 - 7 focus on child health and encourage families to consult health care provider whenever necessary • Session 5 and 6 specifically talk about preventable diseases, their symptoms, and child immunization • Sessions 4 - 7 emphasize exclusive breastfeeding for 6 months, and explain child's nutritional requirements with growing age • These sessions also include homemade recipes for adding semisolid diets for children older than 6 months
Environment for child's development	• No awareness and understanding of child stimulation and play • Families focus on child nutrition and link nutrition only with physical growth	• Sessions 4 - 7 explain stimulation and play, and provide age-appropriate play activities for children • Content on making toys from everyday items to enhance child stimulation and play

[a]PHC: primary health care.

[b]CHW: community health worker.

[c]TBA: traditional birth attendant.

[d]SBA: skilled birth attendant.

Participant mothers and CHWs preferred stories to convey health information. A creative team comprised of a graphic designer, a scriptwriter, and a video producer was engaged to record short docudramas to address the information gaps. Once the graphics and audio-visual clips were ready, we collaborated with the application development team to transfer the entire set into a user-friendly, easy-to-navigate digital package. Called Sehat Ghar, this digital application was installed on a tablet for the CHWs to use in their counseling sessions with the mothers. The cost of development was US $1000 for the 12 videos and US $3000 for the development of the Android application. Upon CHWs' recommendation, a wall calendar for mothers was also developed that displayed the graphics and short messages from the video. In between the two visits, this calendar would remind the mother about the actions she needed to carry out for her and her baby. Input from CHWs and the community was incorporated into these pieces at all stages of the intervention development.

The Sehat Ghar intervention comprised 7 sessions that used 14 live-action videos and a health calendar (Table 1). Delivered by the CHW at the mother's house at monthly intervals, the sessions started from the last trimester of pregnancy until the child was aged 6 months. The content included timely and appropriate messages on maternal health and nutrition, child

health and nutrition, the importance of a skilled birth attendant, immunization, and a positive home environment. With this package, we theorized that CHW-delivered Sehat Ghar content would improve the knowledge and self-efficacy of mothers about the steps that can be ensured at home. Moreover, family practices, including facilitating the mother to go to a nearby public health facility for health care will also improve. The combined result of these practices will be an improvement in mother and child health, and child development.

Deployment of the Program

A 2-day training was offered to all 15 CHWs enrolled in the study. Due to personal or family reasons, 5 dropped out at different time points, and a final set of 10 CHWs completed all steps of training and apprenticeship. The process comprised the main training followed by 1 week of apprentice work in the field and a 1-day training refresher after CHWs had completed apprenticeship in the field. The training curriculum comprised video-based health education content to address gaps in the knowledge about mother and child health, interpersonal communication, and skill development for using the tablet and the Sehat Ghar application.

The CHW-led sessions comprised two live-action videos for discussion. The videos presented a contrasting story through human characters. The first story showed a household where the family did not encourage a mother's healthful nutrition, with the mother and her baby ultimately facing health consequences. The second story presented a family with a similar socioeconomic status but positive thinking and self-belief, dealing with economic problems yet ensuring healthy food for the mother and the baby. After watching the videos together, the CHW discussed similar problems the participant mother and family may be facing and the best ways to jointly solve them. Applying a simplified cognitive behavioral technique (CBT) used in our community-based studies earlier [21], the CHW discussed the practical activities that a mother and family could implement to solve those problems. In a nondidactic manner, the CHW would emphasize interspousal discussions for problem-solving and making the best use of available resources.

Implementation and Testing

Once the CHW started home visits, a field supervisor (also involved in CHW training) regularly conducted field visits to observe the field performance and practices of CHWs. Using a 21-item (yes/no) checklist, the supervisor assessed the CHW's engagement with the family. At the end of each visit, the supervisor appreciated a satisfactory CHW performance and discussed any areas of improvement. Data from the initial 20 visits (2 visits per CHW) were analyzed to carry out the performance ranking of CHWs and draw lessons for collective feedback. On the first follow-up after 3 days, qualitative feedback was also obtained from CHWs about the Sehat Ghar application and use of the digital tablet. In addition, the supervisor also obtained qualitative feedback from mothers about the usefulness of CHW visits and the content of the Sehat Ghar application.

Results

A final cohort of 10 female CHWs with a minimum of 10 years of schooling and who reside in the same community participated in this formative study. The knowledge assessment of these CHWs before and after the training (Table 2) showed significant improvement (mean knowledge score before training: mean 8.00, SD 1.49; mean score post training: mean 11.40, SD 1.43; $P<.001$) after the training. In the posttraining qualitative feedback, CHWs showed confidence in their competency and community acceptance. They appreciated Sehat Ghar, especially its videos. One shared, "As soon as we play the video, the training gets refreshed in our mind. Also, the videos give us words to communicate better." Some CHWs reported that their clients complained that the videos were lengthy. The CHWs improvised by narrating a brief story and going on to the discussion. One CHW shared:

> *There are women who do not seem to have enough time. They are always busy with their children or other work and make excuses when we visit them. I do not show full videos in that case but tell them the main story. Eventually, I think, these women will develop interest.*

During the household visits observed for 20 families (Table 3), the CHWs showed high *fidelity to intervention*. They fulfilled (90% of visits) all the required steps except for praising the mother and family on accomplishments, which were observed during 58% of visits. The *mastery of using tablet* was very good to excellent, as all items received a high score (83% - 100%) in all the visits. Five items in *the completeness of intervention* received good to excellent scores (85% - 92%), while the other 5 received moderate to good (33% - 80%) scores. The items that did not receive an adequate score included using the main picture to build discussion, giving the mother enough time to speak, using the full session content, including videos, and reminding the mother about the practical tasks.

Table . Mean knowledge scores of community health worker before and after the training (N=10).

Question	Before, mean (SD)	After, mean (SD)	*P* values
Q1. How many ANC[a] are required during a pregnancy?	0.50 (0.53)	0.9 (0.31)	.02
Q2. Most women take iron tablets during pregnancy; what is the benefit?	0.70 (0.49)	0.8 (0.42)	.17
Q3. Out of these who is a skilled birth attendant? (more than 1 options allowed)	0.70 (0.49)	1 (0)	.04
Q4. What should be given to a newborn as the first food?	0.80 (0.38)	0.9 (0.31)	.17
Q5. When should we bathe a baby first time after birth?	0.50 (0.53)	0.9 (0.31)	.05
Q6. For how long should a baby receive only mother's milk?	0.70 (0.49)	0.9 (0.31)	.08
Q7. In summer, a 3-mo old baby can be given water with breastfeeding. True or false?	0.50 (0.53)	0.9 (0.31)	.02
Q8. How many times should a child receive vaccination until 15 mo of age?	0.30 (0.48)	0.6 (0.51)	.04
Q9. What effect will vaccination have on a child?	0.90 (0.31)	1 (0)	.17
Q10. A child starts learning immediately after birth.	0.20 (0.42)	0.8 (0.42)	.002
Q11. Special toys are necessary in order to play with a child.	0.80 (0.42)	0.9 (0.31)	.17
Q12. Who is responsible for a child's good upbringing?	0.60 (0.52)	0.8 (0.42)	.08
Q13. Home environment has an effect on a child's physical and cognitive development.	0.80 (0.42)	1 (0)	.08
Total	8.00 (1.49)	11.40 (1.43)	.001

[a]ANC: antenatal checkup.

Table . Community health worker's performance in engaging with family and using the tablet and application during the 20 visits.[a]

Observation		Yes, n (%)	No, n (%)
Fidelity			
	1. Greets the family	20 (100)	0 (0)
	2. Has a friendly interaction with mother/other family members	20 (100)	0 (0)
	3. Asks questions about everyday routine to build the conversation	18 (90)	2 (10)
	4. Talks about previous visit, listens to what mother has to say	18 (90)	2 (8)
	5. Praises the mother/family over tasks they have done.	12 (58)	8 (42)
	6. Gives satisfactory answers to mother/family's questions before moving on to the new content	18 (92)	2 (8)
Mastery of using the tablet			
	7. Can lock/unlock the gadget without difficulty	20 (100)	0 (0)
	8. Is confident about swiping between different screens while using the application	10 (83)	2 (17)
	9. Knows all the key functions well and when to use them. (volume/power/back key)	17 (83)	3 (17)
	10. If the device gets locked while delivering the visit, knows how to turn it back on.	20 (100)	0 (0)
	11. Takes care of the protection of the gadget while using it.	20 (100)	0 (0)
Command over the content			
	12. Picks the appropriate visit according to month of pregnancy/age of child.	18 (92)	2 (8)
	13. Uses the main picture to build a discussion around the topic	4 (20)	16 (80)
	14. Asks the mother's views first regarding the behavior of focus	12 (58)	8 (42)
	15. Pauses in between the videos, asks mother about her opinion	17 (85)	3 (15)
	16. Discusses the problem and its potential solution	18 (92)	2 (8)
	17. Shows complete video	13 (67)	7 (33)
	18. Shows both the videos	18 (92)	2 (8)
	19. Is fluent with the content	17 (85)	3 (15)
	20. Talks about all the pictorial action points	13 (67)	7 (33)
	21. Reminds the mother about practical tasks, marks them on the calendar.	13 (67)	7 (33)

[a]Scoring: excellent >90%, very good 80% - 89%, good 70% - 79%, need improvement <69%.

In a typical visit, the CHWs spent 45 - 60 minutes with the mother and her family. Mothers and their families were happy to have the CHW visit them and discuss the mother and baby's health. "Nobody has ever come to visit us as these women do,

and no one even comes to give us information verbally," said one mother. Mothers appreciated the idea of conveying helpful information through videos and wanted to receive it consistently. They said it increased their knowledge and changed some of their practices. One shared:

> *There were many things we did not know or were doing wrong. I was taking iron tablets with milk, learned the correct method through these visits, and now I take them with water.*

The study team and the CHWs also faced a few difficulties. The absence of a public sector health care system in the study areas narrowed the chances for the team to access the households. To open doors, we engaged 15 health workers but had to work with 10 from start to end, as the rest had to drop out for the aforementioned reasons. Internet connections, required mainly to download the application updates, were sporadic in some field areas. The study team, therefore, timed all the Sehat Ghar updates with monthly meetings the CHWs held at the main office where high-speed internet was available. The length of the videos was a challenge for a few mothers because they had to balance watching the videos with domestic responsibilities. However, the majority were interested in the characters and stories presented in the videos and watched them from start to end.

Discussion

In this study, we developed a digital intervention that volunteer CHWs could use to deliver health education with basic training and some supervision. The exploratory part of this study revealed that families living in underserved areas unattended by PHC services were unaware of basic health promotion and disease prevention. Women relied on traditional birth attendants for antenatal checkups and went to the hospital only at the time of delivery, with no or minimal postnatal checkups. Newborn care was marginal as was the knowledge about a growing child's health and nutrition. The home environment offered mixed conditions for child development. With the help of minimally trained volunteers from the community, mothers received infotainment-based health education. The volunteers and their audience mothers liked the home visits and the content of the intervention and wished to engage in this activity regularly.

Using videos as a tool to deliver counseling sessions had several advantages. First, the story format was interesting both for the mother and the health worker. Moreover, the characters (shown in an environment similar to the participant family) were taking actions about which health workers would ask questions warranting the mother's attention. Second, the same videos were being used by all health workers ensuring adequate dose and fidelity [32]. This complete and acceptable delivery of interventions has been reported as a challenge when print materials are used along with the oral delivery of the content [33]. Third, the same videos could be used by thousands of lady health workers of the national program [34] after successful proof of effectiveness.

In contrast to some of the published studies in which such sessions turned into a passive watching of the video [35], our intervention used CBT-based principles that provided a structure to the session, with the video acting as a supporting tool. This meant that both the health worker and the mother must actively watch the video in anticipation of the questions that the health worker would ask, and the mother would respond about the situations faced by the characters in the video. Each discussion finished with both the health worker and the mother or other family members discussing the relevance of the content with their situation, identifying their problems, and working on joint solutions as reported in earlier studies [36].

Technology-based interventions like SMS text message reminders via mobile phones, robocalls, and social media have been used in the past, with their top-down communication and delivery from a distance limiting their effectiveness [37,38]. Studies assessing other behavior change strategies, like messaging campaigns through mass media, have found that they may not be effective when used in isolation [39]. Combining media and technology with face-to-face communication works better [40]. However, health content delivered through face-to-face communication by workers trained in a cascade setting dilutes the intervention with each step of trickle-down and must be addressed by a "cascade-plus" approach [41]. We embedded our digital application in a cascade-plus training of CHWs in which they were provided more than one job aid and multiple learning opportunities to discuss and improve their skills.

Digital interventions for CHW programs have received criticism because of their costs, the need for internet coverage, and intense training requirements [12]. However, our experience was different. We spent a total of US $4000 on this digital intervention—a one-time cost on an application and its videos that can be used by thousands of CHWs. Similarly, Android phones (US $100 in the local market) are owned by a large majority in the country and this application can be adapted for mobile phones. The common availability of smartphones helped in training the CHWs because most of them already knew the operating functions of the device. The lack of high-speed internet in the community did not pose a substantial challenge as the Sehat Ghar application was not dependent on installing updates while in the field or uploading data from the field.

In addition to videos, the identification and training of volunteer women and preparing them for the basic CHW role is also a critical element of our intervention. Such volunteers can be particularly relevant to settings where women have nominal education, families are resource poor, and deeply embedded cultural practices impede the adoption of healthful behaviors. In a community without the coverage of PHC, we were able to identify and train health volunteers who learned using a video-based digital application and then provided health education to mothers and families to facilitate a positive change in their health behaviors. This strategy can be useful in emergency and disaster situations, where infrastructure is badly damaged and essential health services have to be restored quickly.

Though this study was conducted before the pandemic, it has some insights applicable to outbreak situations as well. The video-based health content could be uploaded to other media

for a consumer interested in watching stories to learn about her health. The same could be done by adding video content on the prevention of COVID-19 for an internet user (or a health worker) to download amid the lockdowns. Two similar "distance-compliant" strategic communications in Pakistan (ie, COVID-19 prevention messages used as caller ringtones on mobile phones, and the 24-hour national call-in helpline 1166) proved to be highly effective during the pandemic [42]. This implies that a digital health app already present on a mobile phone can be quickly updated with new information, including videos, in outbreak situations.

Some limitations are worth mentioning for a better interpretation of our findings. We wanted to conduct as many interviews and FGDs as required to reach a theoretical saturation point but had to limit our sample given the access issues. The high improvement in posttraining knowledge scores could be due to a smaller set of trainees who received good attention from the trainers. Moreover, they had less health knowledge previously, which could improve quickly. However, a change in knowledge does not guarantee a change in behaviors—the change in CHW practices observed over time will be a more meaningful outcome. The observations too have their limitation because of social desirability bias. However, we conducted these observations as part of "supportive supervision" in which the CHW knew that she would be helped and not scolded if she did not remember something and asked about it.

While we look forward to the implementation of Sehat Ghar among the 250 families to provide important insights for upscaling, we also envisage a few challenges for large-scale real-world implementation. Adopting a new approach of counseling that we developed using CBT principles may be resisted because the countrywide lady health worker program will have to plan new training. Moreover, after adoption, these visits will need monitoring and supportive supervision across the country to help the workers and avoid passive use of the videos, as reported by other studies [36]. However, the application's development was a one-time cost, and the application or only the videos can be used after downloading on thousands of digital devices.

Based on the results from this pilot phase, we concluded that developing an easy-to-use digital application like Sehat Ghar is possible with minimal resources. Its use by minimally trained volunteers and acceptance by the community make it feasible. Its systematic rollout can contribute to improving health behaviors, including health care use. Mothers appreciate interesting content like video stories and are willing to incorporate the information into their daily practices. Lastly, applying participatory approaches during the formative phase of such interventions is critical for their long-term adoption by the community and CHW programs.

Acknowledgments

The study was funded by Grand Challenges Canada through a seed grant (ST-POC-1808-17445).

Authors' Contributions

ZUH conceptualized the study. ZUH, AN, and NuAP oversaw the data collection and analysis. DZ and MS helped in the final analysis and writing of different sections. All authors read and approved the final draft.

Conflicts of Interest

None declared.

References

1. Ghebreyesus TA, Fore H, Birtanov Y, Jakab Z. Primary health care for the 21st century, universal health coverage, and the Sustainable Development Goals. Lancet 2018 Oct 20;392(10156):1371-1372. [doi: 10.1016/S0140-6736(18)32556-X] [Medline: 30343843]
2. de Brito Lima Prado NM, Rossi TRA, Chaves SCL, et al. The international response of primary health care to COVID-19: document analysis in selected countries. Cad Saude Publica 2020 Nov 20;36(12):e00183820. [doi: 10.1590/0102-311X00183820] [Medline: 33237251]
3. Sacks E, Schleiff M, Were M, Chowdhury AM, Perry HB. Communities, universal health coverage and primary health care. Bull World Health Organ 2020 Nov 1;98(11):773-780. [doi: 10.2471/BLT.20.252445] [Medline: 33177774]
4. Ludwick T, Morgan A, Kane S, Kelaher M, McPake B. The distinctive roles of urban community health workers in low- and middle-income countries: a scoping review of the literature. Health Policy Plan 2020 Oct 1;35(8):1039-1052. [doi: 10.1093/heapol/czaa049] [Medline: 32494801]
5. Hartzler AL, Tuzzio L, Hsu C, Wagner EH. Roles and functions of community health workers in primary care. Ann Fam Med 2018 May;16(3):240-245. [doi: 10.1370/afm.2208] [Medline: 29760028]
6. Babagoli MA, Nieto-Martínez R, González-Rivas JP, Sivaramakrishnan K, Mechanick JI. Roles for community health workers in diabetes prevention and management in low- and middle-income countries. Cad Saude Publica 2021 Oct 29;37(10):e00287120. [doi: 10.1590/0102-311X00287120] [Medline: 34730688]

7. Olaniran A, Madaj B, Bar-Zev S, van den Broek N. The roles of community health workers who provide maternal and newborn health services: case studies from Africa and Asia. BMJ Glob Health 2019 Aug 10;4(4):e001388. [doi: 10.1136/bmjgh-2019-001388] [Medline: 31478012]
8. Scott K, Beckham SW, Gross M, et al. What do we know about community-based health worker programs? A systematic review of existing reviews on community health workers. Hum Resour Health 2018 Aug 16;16(1):39. [doi: 10.1186/s12960-018-0304-x] [Medline: 30115074]
9. Cometto G, Ford N, Pfaffman-Zambruni J, et al. Health policy and system support to optimise community health worker programmes: an abridged WHO guideline. Lancet Glob Health 2018 Dec;6(12):e1397-e1404. [doi: 10.1016/S2214-109X(18)30482-0] [Medline: 30430994]
10. Amoakoh-Coleman M, Borgstein ABJ, Sondaal SF, et al. Effectiveness of mHealth interventions targeting health care workers to improve pregnancy outcomes in low- and middle-income countries: a systematic review. J Med Internet Res 2016 Aug 19;18(8):e226. [doi: 10.2196/jmir.5533] [Medline: 27543152]
11. Mahmood H, Mckinstry B, Luz S, et al. Community health worker-based mobile health (mHealth) approaches for improving management and caregiver knowledge of common childhood infections: a systematic review. J Glob Health 2020 Dec;10(2):020438. [doi: 10.7189/jogh.10.020438] [Medline: 33437462]
12. White A, Thomas DSK, Ezeanochie N, Bull S. Health worker mHealth utilization: a systematic review. Comput Inform Nurs 2016 May;34(5):206-213. [doi: 10.1097/CIN.0000000000000231] [Medline: 26955009]
13. Feroz AS, Khoja A, Saleem S. Equipping community health workers with digital tools for pandemic response in LMICs. Arch Public Health 2021 Jan 4;79(1):1. [doi: 10.1186/s13690-020-00513-z] [Medline: 33390163]
14. Announcement of results of 7th population and housing census-2023 'The Digital Census'. Pakistan Bureau of Statistics. URL: https://www.pbs.gov.pk/content/announcement-results-7th-population-and-housing-census-2023-digital-census [accessed 2024-06-21]
15. Population living in slums (% of urban population) - Pakistan. World Bank Group: Data. URL: https://data.worldbank.org/indicator/EN.POP.SLUM.UR.ZS?locations=PK [accessed 2024-01-17]
16. National Institute of Population Studies. Pakistan demographic and health survey 2017-18. The DHS Program. 2019 Jan. URL: https://dhsprogram.com/pubs/pdf/FR354/FR354.pdf [accessed 2024-07-15]
17. Rehman A, Shaikh BT, Ronis KA. Health care seeking patterns and out of pocket payments for children under five years of age living in Katchi Abadis (slums), in Islamabad, Pakistan. Int J Equity Health 2014 Apr 16;13(1):30. [doi: 10.1186/1475-9276-13-30] [Medline: 24739395]
18. Yaya S, Anjorin SS, Adedini SA. Disparities in pregnancy-related deaths: spatial and Bayesian network analyses of maternal mortality ratio in 54 African countries. BMJ Glob Health 2021 Feb;6(2):e004233. [doi: 10.1136/bmjgh-2020-004233] [Medline: 33619040]
19. Peven K, Bick D, Purssell E, Rotevatn TA, Nielsen JH, Taylor C. Evaluating implementation strategies for essential newborn care interventions in low- and low middle-income countries: a systematic review. Health Policy Plan 2021 Nov 1;35(Supplement_2):ii47-ii65. [doi: 10.1093/heapol/czaa122] [Medline: 33156939]
20. Bhutta ZA, Das JK, Rizvi A, et al. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet 2013 Aug 3;382(9890):452-477. [doi: 10.1016/S0140-6736(13)60996-4] [Medline: 23746776]
21. Rahman A, Haq Z, Sikander S, Ahmad I, Ahmad M, Hafeez A. Using cognitive-behavioural techniques to improve exclusive breastfeeding in a low-literacy disadvantaged population. Matern Child Nutr 2012 Jan;8(1):57-71. [doi: 10.1111/j.1740-8709.2011.00362.x] [Medline: 22066882]
22. Liu Y, Jiang Y. Urban growth sustainability of Islamabad, Pakistan, over the last 3 decades: a perspective based on object-based backdating change detection. GeoJournal 2021 Oct;86(5):2035-2055. [doi: 10.1007/s10708-020-10172-w]
23. Sawas A, Mustafa D, Anwar NH, Iqtidar H, Viqar S. Urbanization, gender and violence in Rawalpindi and Islamabad: a scoping study. ResearchGate. URL: https://www.researchgate.net/publication/278670851_Urbanization_Gender_and_Violence_in_Rawalpindi_and_Islamabad_A_Scoping_Study [accessed 2024-07-15]
24. Shafqat R, Marinova D, Khan S. Placemaking in informal settlements: the case of France colony, Islamabad, Pakistan. Urban Sci 2021 Jun 20;5(2):49. [doi: 10.3390/urbansci5020049]
25. August F, Pembe AB, Mpembeni R, Axemo P, Darj E. Effectiveness of the home based life saving skills training by community health workers on knowledge of danger signs, birth preparedness, complication readiness and facility delivery, among women in rural Tanzania. BMC Pregnancy Childbirth 2016 Jun 2;16(1):129. [doi: 10.1186/s12884-016-0916-x] [Medline: 27251052]
26. Hafeez A, Mohamud BK, Shiekh MR, Shah SAI, Jooma R. Lady health workers programme in Pakistan: challenges, achievements and the way forward. J Pak Med Assoc 2011 Mar;61(3):210-215. [Medline: 21465929]
27. Patton MQ. Qualitative Research & Evaluation Methods, 3rd edition: SAGE Publications; 2002.
28. Cresswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 3rd edition: SAGE Publications; 2013.
29. Miles M, Huberman M. Qualitative Data Analysis: An Expanded Sourcebook, 2nd edition: SAGE Publications; 1994.

30. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020 Jan;21(1):37-48. [doi: 10.1177/1524839919850557] [Medline: 31131633]
31. Blynn E, Harris E, Wendland M, et al. Integrating human-centered design to advance global health: lessons from 3 programs. Glob Health Sci Pract 2021 Nov 29;9(Suppl 2):S261-S273. [doi: 10.9745/GHSP-D-21-00279] [Medline: 34845049]
32. Saunders RP. Implementation Monitoring and Process Evaluation, 1st edition: SAGE Publications, Inc; 2016. [doi: 10.4135/9781071878736]
33. Giguère A, Zomahoun HTV, Carmichael PH, et al. Printed educational materials: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2020 Jul 31;8(8):CD004398. [doi: 10.1002/14651858.CD004398.pub4] [Medline: 32748975]
34. Wazir MS, Shaikh BT, Ahmed A. National program for family planning and primary health care Pakistan: a SWOT analysis. Reprod Health 2013 Nov 22;10(1):60. [doi: 10.1186/1742-4755-10-60] [Medline: 24268037]
35. Greuel M, Sy F, Bärnighausen T, et al. Community health worker use of smart devices for health promotion: scoping review. JMIR Mhealth Uhealth 2023 Feb 22;11:e42023. [doi: 10.2196/42023] [Medline: 36811947]
36. Sikander S, Maselko J, Zafar S, et al. Cognitive-behavioral counseling for exclusive breastfeeding in rural pediatrics: a cluster RCT. Pediatrics 2015 Feb;135(2):e424-e431. [doi: 10.1542/peds.2014-1628] [Medline: 25583916]
37. Clar C, Dyakova M, Curtis K, et al. Just telling and selling: current limitations in the use of digital media in public health: a scoping review. Public Health 2014 Dec;128(12):1066-1075. [doi: 10.1016/j.puhe.2014.09.009] [Medline: 25443388]
38. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
39. Sarrassat S, Meda N, Ouedraogo M, et al. Behavior change after 20 months of a radio campaign addressing key lifesaving family behaviors for child survival: midline results from a cluster randomized trial in rural Burkina Faso. Glob Health Sci Pract 2015 Nov 3;3(4):557-576. [doi: 10.9745/GHSP-D-15-00153] [Medline: 26681704]
40. Zhang D, Jin L, Liang D, et al. Assessing feasibility of an early childhood intervention using mobile phones among low-income mothers of newborns: qualitative interview study. JMIR Form Res 2020 May 28;4(5):e17179. [doi: 10.2196/17179] [Medline: 32463374]
41. Öztek Z, Kydd A. A new training approach for vaccinators: cascade plus training. Turk J Public Health 2022 Apr;20(1):164-176. [doi: 10.20518/tjph.1017386]
42. Haq ZU, Mirza Z, Oyewale TO, Sultan F. Leaving no one behind: Pakistan's risk communication and community engagement during COVID-19. J Glob Health 2021 Jul 31;11:03091. [doi: 10.7189/jogh.11.03091] [Medline: 34386212]

Abbreviations

CBT: cognitive behavioral technique
CHW: community health worker
FGD: focus group discussion
HCP: health care provider
IDI: in-depth interview
PHC: primary health care

Edited by E Meinert; submitted 15.04.23; peer-reviewed by Anonymous, J Johnston; revised version received 12.04.24; accepted 29.04.24; published 31.07.24.

Please cite as:
Haq ZU, Naeem A, Zaeem D, Sohail M, Pervaiz NUA
Development of a Digital Platform to Promote Mother and Child Health in Underserved Areas of a Lower-Middle-Income Country: Mixed Methods Formative Study
JMIRx Med 2024;5:e48213
URL: https://xmed.jmir.org/2024/1/e48213
doi:10.2196/48213

Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers

Jaime Perales-Puchalt[1], MPH, PhD; Jill Peltzer[2], PhD, APRN-CNS; Monica Fracachan-Cabrera[2], MD; G Adriana Perez[3], PhD, ANP-BC; Mariana Ramírez[2], LCSW; K Allen Greiner[2], MD; Jeffrey Murray Burns[1], MD

1

2

3

Corresponding Author:
Jaime Perales-Puchalt, MPH, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/42211

Companion article: https://www.medrxiv.org/content/10.1101/2022.05.25.22275517v2

Companion article: https://med.jmirx.org/2024/1/e56444

Companion article: https://med.jmirx.org/2024/1/e56443

Companion article: https://med.jmirx.org/2024/1/e57160

Companion article: https://med.jmirx.org/2024/1/e56439

Abstract

Background: Latino individuals experience disparities in the care of Alzheimer disease and related dementias (ADRD) and have disproportionately high COVID-19 infection and death outcomes.

Objective: We aimed to gain an in-depth understanding of the impact of the COVID-19 pandemic among Latino families with ADRD in the United States.

Methods: This was a qualitative study of 21 informal caregivers of Latino individuals with ADRD and 23 primary care providers who serve Latino patients. We recruited participants nationwide using convenience and snowball sampling methods and conducted remote interviews in English and Spanish. We organized the transcripts for qualitative review to identify codes and themes, using a pragmatic approach, a qualitative description methodology, and thematic analysis methods.

Results: Qualitative analysis of transcripts revealed eight themes, including (1) the pandemic influenced mental and emotional health; (2) the pandemic impacted physical domains of health; (3) caregivers and care recipients lost access to engaging activities during the confinement; (4) the pandemic impacted Latino caregivers' working situation; (5) the pandemic impacted health care and community care systems; (6) health care and community care systems took measures to reduce the impact of the pandemic; (7) Latino families experienced barriers to remote communication during the pandemic; and (8) caregiver social support was critical for reducing social isolation and its sequalae.

Conclusions: Latino families with ADRD experienced similar but also unique impacts compared to those reported in the general population. Unique impacts may result from Latino individuals' underserved status in the United States, commonly held cultural values, and their intersectionality with ADRD-related disability. Family caregiver social support was crucial during this time of adversity. These findings suggest the need for more equitable access, culturally appropriate and trustworthy content and delivery of health care and community services, as well as stronger financial and social supports for family caregivers.

(JMIRx Med 2024;5:e42211) doi:10.2196/42211

KEYWORDS

Latino; dementia; caregiving; COVID-19; Alzheimer's disease; health disparity; qualitative research; transcript analysis; health inequality; minority population; epidemiology; peer support; social services; health care; Alzheimer's; minority; qualitative; interview; caregiver; primary care; impact; resilience; disparity; outcome; Alzheimer disease; Alzheimer

Introduction

Alzheimer disease and related dementias (ADRD) have a devastating impact on Latino families. Compared to non-Latino White individuals, Latino individuals experience disparities in ADRD risk, detection, treatment, and care. While Latino individuals are 1.5 times more likely to have ADRD [1], they are 1.4 times less likely to be diagnosed [2]. Latino individuals are diagnosed with an 8-month delay [3], are underprescribed cholinesterase inhibitors [4], and underuse ADRD support services [5]. These disparities are further compounded by Latino individuals' lower ADRD knowledge and inclusion in ADRD research [6-8].

Latino families have also been disproportionately impacted by the COVID-19 pandemic. As of January of 2022, among individuals aged 65 years and older, Latino individuals represent 8% of the US population but 13% of COVID-19 cases [9]. Latino individuals of any age are also more likely to be infected with COVID-19, representing 18% of the US population but 25% of those infected. When controlling for age, the percentage of COVID-19 deaths is also higher among Latino individuals. For example, Latino individuals represent 14% of the deaths among those aged 65 years and older, which is nearly twice as much as their representation in the US population aged 65 years and older [9]. Latino individuals are also less likely to be vaccinated against COVID-19, representing only 12% of those initiating vaccination [10]. Some factors contributing to these disparities include having a job considered essential, living in a segregated geographic area, living in overcrowded households, limited English proficiency, and reduced use of preventive behaviors to avoid COVID-19 infection [11].

Despite the growing literature on the impact of COVID-19 among Latino individuals, little is known about the impact of COVID-19 on Latino families with ADRD. Studying COVID-19's impact on Latino families with ADRD is important given that Latino individuals are populations with disparities in both ADRD and COVID-19, both of which might synergistically impact their health [9,10,12]. A study in California that included 8 Latino individuals with cognitive impairment found that the biggest impact of COVID-19 on these and other underserved communities included the fear generated by the pandemic, distress stemming from feeling extremely isolated, and receiving inaccurate information about COVID-19 from different sources [13]. The study also reported some strategies participants used for coping during the pandemic (eg, mask wearing and remote communication) and the importance of access to essential resources such as friends, the church, and local programs. Although this work provided important insights on the impact and resilience factors for diverse populations with cognitive impairment, it is crucial to increase the Latino representation for a more detailed understanding. It is also important to listen to the perspectives of family caregivers and primary care providers (PCPs), who may provide a different point of view and allow triangulation. Given the central role of family in Latino culture [14] and PCPs in the health care system [15], the perspectives of family members and PCPs can provide privileged insight into the experiences of persons with ADRD, irrespective of their level of cognitive impairment, their family, and their interaction with the health care system.

The aim of this study was to gain an in-depth understanding of the impact of the COVID-19 pandemic among Latino families with ADRD. To achieve this goal, we interviewed family caregivers and PCPs who serve Latino individuals with ADRD across the United States. Findings can inform actions to help Latino families with ADRD remain safe while maintaining a good quality of life during this ongoing pandemic as well as future ones.

Methods

Assessment

This study used a qualitative descriptive design. Qualitative studies rely on text data and aim to understand the meaning of human action [16]. The current COVID-19–focused analysis is part of a broader study that planned to identify what ADRD care services are offered in primary care and how they are delivered across a variety of settings. This study was informed by the *National Institute on Minority Health and Health Disparities Research Framework*, which considers the complex and multifaceted nature of minority health and health disparities [17]. This framework includes different domains of influence (biological, behavioral, physical or built environment, sociocultural environment, and health care system), as well as different levels of influence (individual, interpersonal, community, and societal) within these domains.

Ethical Considerations

The data collected were stripped of all identifying information and labeled with a code. The code can be linked to the individual's identity, but the key linking the code to the identifying information is held separately from the research data. The University of Kansas Medical Center Institutional Review Board approved this project (STUDY00145615). Before the interview, all participants completed an informed written consent on the web either via their computers, tablets, or phones. The research team compensated participant's time by posting them a US $40 gift card.

Sample

We recruited 2 groups of participants: (1) family caregivers of people living with ADRD and (2) PCPs. Inclusion criteria for caregivers included being aged 18 years or older, identifying as a close friend or relative of a Latino person with ADRD diagnosed by a health care provider or research study, providing or having recently provided care to them at least once a week in person or via phone, being proficient in English or Spanish, living in the United States, and being willing to participate in

the study. Inclusion criteria for PCPs included being a medical doctor, doctor in osteopathic medicine, nurse practitioner, or physician assistant who currently or recently provided primary care services to Latino families with ADRD in the United States.

Caregivers were recruited via convenience sampling from diverse sources, including a research registry, clinics, and community ADRD patient lists, internet and newspaper advertisements, and connections with community partners and research team members. PCPs were initially recruited using snowball sampling techniques: first by contacting connections of the research team and later asking interviewed PCPs for referral of other PCPs.

We used purely qualitative semistructured interviews, which allow for the comparison between participants while allowing spontaneous exploration of topics relevant to unique participants [18]. The interviews took place between November 2020 and April 2021. All but 2 interviews were conducted via secure videoconference. The other 2 interviews were conducted over the phone. The interviewer used 2 audio recorders to reduce the risk of data collection failure and took notes at the end of each interview to summarize the findings. Questions were not provided to participants before the interview. No one else besides the interviewer and the participant was present during interviews, although in a few instances, participants had to pause the interview briefly to address issues with their loved ones (caregivers) or colleagues (PCPs).

The process for each interview was similar: JP-P interviewed all participants. The interview started with a short conversation aimed at developing rapport and explaining the main goals of the interview. The first questions asked about the basic characteristics of the participants. Core questions of the survey asked about participants' experience with primary care clinics (caregivers) and serving Latino individuals with ADRD (PCPs). Unless participants had mentioned it spontaneously, the interviewer asked participants halfway through the interview how the COVID-19 pandemic had impacted them, their care recipient (caregivers), and the Latino families with ADRD they serve (PCPs; see Multimedia Appendix 1 for the interview guides). The interview questions had been previously pilot-tested within the team. The interviewer audiotaped all interviews, which were designed to last 45-60 minutes, including the COVID-19 questions. Interviews were conducted in English or Spanish, and those with caregivers included simple language to account for different levels of literacy. A professional team transcribed all interviews and the interviewer reviewed them for accuracy.

To bring rigor and validity to the research process, the interviewer used active listening techniques during the interview aimed at confirming the information shared by participants. The potentially lower theoretical ADRD expertise by family members and PCPs could create a perceived power differential with the interviewer. In fact, some family members and PCPs expressed worry about the interviewer potentially testing their ADRD knowledge. To reduce this perception of power differential, the interviewer emphasized that the interviews would ask about their experiences and that they were the experts in those experiences. JP-P already knew some of the family caregivers via his research but knew none of the PCPs. To balance how much participants knew about him, he introduced himself, his background, previous work, and research goals the first time he contacted them and at the beginning of the interview. Both coders (JP-P and MF-C) had previous coding experience. The fact that only 1 coder conducted the interviews and that they both had different backgrounds (JP-P: PhD, man, faculty, and psychology and public health education vs MF-C: MD, woman, research associate, and medical education) allowed different perspectives.

Data Analysis

We organized the transcripts and notes for qualitative review, using a pragmatic approach and thematic analysis methods, which emphasize identifying, analyzing, and interpreting patterns of meaning within qualitative data [19,20]. We organized the data in Dedoose [21]. JP-P identified relevant sociodemographic, ADRD relationship, and clinical service data in the interviews and summarized it into descriptive statistics. JP-P and MF-C independently read the interviews and notes once initially to familiarize themselves with the data, coded the content of the text by identifying codes and themes, and resolved coding disagreements through discussion and consensus.

Results

Characteristics of the Sample

Considering the time-sensitive nature of this work, we decided to stop recruitment when data saturation was achieved at 44 participants. Table 1 shows the characteristics of the 21 family caregivers and 23 PCPs. Most caregivers were women (19/21, 90%), younger than 66 years (16/21, 76%), and children of the care recipient (15/21, 71%), most of whom had Alzheimer disease (14/21, 67%). All but 1 (20/21, 95%) participant identified as Latino of diverse origins; 6 (29%) were born in the United States; and all lived in urban regions in the Midwest (15/21, 71%), the Northeast (5/21, 24%), and the South (1/21, 5%). Interviews were conducted in Spanish with 13 (62%) caregivers, 14 (67%) of them reported good English proficiency, and 8 (38%) reported that their care recipient had good English proficiency.

Table . Characteristics of the sample.

Characteristics		Family caregivers (n=21)	Primary care providers (n=23)
Women, n (%)		19 (90)	13 (57)
Age group (years), n (%)			
	33-65	16 (76)	N/A[a]
	66+	5 (24)	N/A
Ethnoracial group, n (%)			
	Latino	20 (95)	10 (43)
	Non-Latino Asian	0 (0)	1 (4)
	Non-Latino Black	0 (0)	2 (9)
	Non-Latino White	1 (5)	10 (43)
Subgroup among Latino participants, n (%)[b]			
	Caribbean	3 (14)	3 (13)
	Central American	4 (19)	1 (4)
	Mexican	10 (48)	2 (9)
	South American	3 (14)	3 (13)
US born, n (%)		6 (29)	12 (52)
Live in an urban setting, n (%)		21 (100)	18 (78)
Interviewed in Spanish, n (%)		13 (62)	5 (22)
Region, n (%)			
	Midwest	15 (71)	15 (65)
	Northeast	5 (24)	2 (9)
	West	0 (0)	4 (17)
	South	1 (5)	0 (0)
	Puerto Rico	0 (5)	2 (9)
Good English proficiency, n (%)		14 (67)	N/A
Can provide services in Spanish, n (%)		N/A	12 (52)
Relation to care recipient, n (%)			
	Child	15 (71)	N/A
	Spouse	5 (24)	N/A
	Friend	1 (5)	N/A
Diagnosis of care recipient, n (%)			
	Alzheimer disease	14 (67)	N/A
	ADRD[c] (unspecified)	2 (10)	N/A
	Mild cognitive impairment	2 (10)	N/A
	Early onset Alzheimer disease	1 (5)	N/A
	Early onset Alzheimer disease and frontotemporal dementia	1 (5)	N/A
	Parkinson dementia	1 (5)	N/A
More than 1 care recipient, n (%)		1 (5)	N/A
Care recipient has good English proficiency, n (%)		8 (38)	N/A
Type of provider, n (%)			
	Medical doctor	N/A	15 (65)
	Nurse practitioner	N/A	7 (30)

Characteristics		Family caregivers (n=21)	Primary care providers (n=23)
	Physician assistant	N/A	1 (4)
Type of clinic, n (%)			
	Private	N/A	11 (48)
	Safety net and federally qualified	N/A	8 (35)
	Academic	N/A	4 (17)
Percent of Latino patients, n (%)			
	Less than 29%	N/A	7 (30)
	30%-74%	N/A	7 (30)
	75%+	N/A	9 (39)
Total minutes of recording, range		25-61	35-73

[a]N/A: not applicable.

[b]Only 9 Latino primary care providers shared subgroup information.

[c]ADRD: Alzheimer disease and related dementias.

More than half of PCP participants were women (13/23, 57%) and US born (12/23, 52%). A total of 10 (43%) PCPs identified as Latino of diverse origins and another 10 (43%) identified as non-Latino White. A total of 18 (78%) PCPs lived in urban settings, with the Midwest being the most common region (15/23, 65%). A total of 12 (52%) PCPs reported being able to communicate with their patients in Spanish. In all, 15 (65%) PCPs were medical doctors and 7 (30%) were nurse practitioners; the types of clinics they practiced in varied, along with a wide distribution of the percentage of Latino patients served.

Themes

We developed 8 themes from the analysis. Table 2 shows the themes and codes identified in the qualitative analyses.

Table . Themes and related codes.

Theme	Code
1. The pandemic influenced mental and emotional health	• 1.1. Fear of infection • 1.2. Uncertainty • 1.3. Distress • 1.4. Depression
2. The pandemic impacted physical domains of health	• 2.1. Accelerated deterioration • 2.2. Poor nutrition • 2.3. Death
3. Caregivers and care recipients lost access to engaging activities during the confinement	• 3.1. Isolation • 3.2. Deprivation
4. The pandemic impacted Latino caregivers' working situation	• 4.1. Avoiding exposure to infection • 4.2. Disrupted work • 4.3. Increased work burden
5. The pandemic impacted health care and community care systems	• 5.1. Health care systems • 5.2. Community care systems
6. Health care and community care systems took measures to reduce the impact of the pandemic	• 6.1. Home health • 6.2. Remote communication
7. Latino families experienced barriers to remote communication during the pandemic	• 7.1. Low trust • 7.2. Insufficient skills and diminished abilities • 7.3. Inadequate support
8. Caregiver social support was critical for reducing social isolation and its sequalae	• 8.1. Protecting from COVID-19 infection • 8.2. Reducing psychosocial impact • 8.3. Reducing health care and community care system impact

Theme 1. The Pandemic Influenced Mental and Emotional Health

The mental and emotional health domains described by caregivers and PCPs included fear of infection, uncertainty, distress, and depression. This was the most reported impact by all interviewees.

1.1. Fear of Infection

The fear of COVID-19 infection for the care recipient was reported by most of the caregivers and PCPs. While some caregivers mentioned that they or their care recipients had not been infected, they were fearful about the risk of COVID-19 to the care recipient, who was at risk for significant disease sequalae. Other caregivers or their care recipient were experiencing COVID-19 at the time of the interview. PCPs explained that this fear was especially felt by minoritized individuals including Latino individuals, given the news that they were getting infected and dying at disproportionate rates compared to non-Latino White individuals. These 2 example quotations summarize this fear:

> *Most ADRD patients are old and this is the population at the highest risk, and they are even more scared.*

> *They do not want to come in. The news said very early that ethnic minorities were getting more severe disease that they just stopped coming in and it is really hard to get a hold of them.*

1.2. Uncertainty

According to both caregivers and PCPs, caregivers and care recipients had feelings of uncertainty. There was uncertainty about when they would be eligible to get the COVID-19 vaccine, when the caregiver would get some respite by being able to take the care recipient back to the senior center, confusion about what was happening due to cognitive impairment, and general uncertainty about their lives. One caregiver defined their lives during the pandemic succinctly as follows:

> *Everything has been in limbo with COVID.*

1.3. Distress

Heightened fear and daily uncertainty led to increased levels of recurring distress. On top of pandemic-related distress, caregivers were unable to access caregiver support services and the respite they needed to manage their own health. They also lacked access to the care residences needed to provide the safest care for the care recipients. This distress added on top of the other stresses Latino families often experience, as this caregiver explained:

> *I was recently told I had cancer. Now I need to get some further testing. This affected me a lot too. It's also been hard to work, take care of the house, my kids…because when it's not one problem it's another. It's been really hard…so much stress.*

1.4. Depression

Caregivers and PCPs highlighted the frequency and severity of depressed mood among caregivers and care recipients, especially during the lockdown due to the lack of social support and increased social isolation. Two PCPs noted the following:

> *All of my ADRD patients are sad…they are isolated, by themselves.*

> *The social isolation of COVID has caused a significant level of depression.*

A caregiver also shared the following:

> *I'm very sad, depressed, neurotic…I've no patience…very little patience. To me, life has changed a huge deal…360 degrees.*

Theme 2. The Pandemic Impacted Physical Domains of Health

The physical domains described by caregivers and PCPs included accelerated deterioration, poor nutrition, and death.

2.1. Accelerated Deterioration

Caregivers and PCPs noticed accelerated cognitive and physical deterioration. One of the caregivers had to move their loved one to a long-term care facility in the midst of the pandemic and saw a rapid decline in cognition. Another PCP reported how the lockdown resulted in reduced access to treatment among some patients, leading to accelerated cognitive and physical decline. The caregivers and PCPs attributed decline to social isolation and a lack of engagement in activities that affected mood, such as depression, sadness, and apathy. A caregiver explained her mother's memory decline during lockdown as follows:

> *I was thinking her memory and abilities declining all the time she was locked down here, because of the cold weather and COVID might have affected her more.*

2.2. Poor Nutrition

Food access was a critical issue, leading to poor nutrition that contributed to accelerated declines in health. Reasons for poor food access included delays in the mailing system and financial insecurity. These barriers interacted with care recipients' impaired abilities and male caregivers' lack of cooking skills and resulted in a frequent use of unhealthier options due to their accessibility. For example, a male caregiver explained that their access to healthy food was impacted because his spouse could not cook anymore due to ADRD:

> *…and my cooking abilities are not all that great…so fast food has been our life for the last year maybe even more because she has not been able to cook for some time.*

2.3. Death

Participants reported death due to COVID-19 among their Latino families. For example, a PCP explained that several of his Latino patients died during their yearly visit to their home country—deaths that would not have occurred prior to the pandemic. A caregiver explained how a relative with ADRD died shortly after being moved to a residential facility:

> *He did not live too long after that, he literally stopped eating…so eventually he was put on Morphine because he was in a lot of pain from multiple things and that is how he passed.*

Theme 3. Caregivers and Care Recipients Lost Access to Engaging Activities During the Confinement

Even though participants understood the need for the lockdown, they frequently described how it contributed to social isolation and deprivation, which, in turn, reduced engagement in activities and led to depression and apathy.

3.1. Isolation

The pandemic paralyzed or slowed down operations in clinics, church services, caregiver support activities, senior centers, and residences. Additionally, PCPs, home health assistants, residence staff, and family members were infected. In fact, a caregiver's former PCP who spoke Spanish had died of COVID-19 and was replaced by a non-Spanish speaker. One of the PCPs mentioned the following:

> *I think it is really a negative impact and really socially isolating a lot of people, which as a PCP always makes me worried because of how much stress that adds to people.*

3.2. Deprivation

Caregivers also reported that their care recipient and themselves lost access to engaging activities during the confinement, such as visiting with their peers, going out to eat, and attending social events at the local senior center. A caregiver summarized this feeling as follows:

> *We have been so stuck with this life of COVID not being able to do anything or go anywhere.*

Theme 4. The Pandemic Impacted Latino Caregivers' Working Situation

The pandemic's impact on Latino caregivers' working situation included avoiding exposure to infection, disrupted work, and increased work burden.

4.1. Avoiding Exposure to Infection

Families had to make difficult choices between exposing frail, older members to COVID-19 or losing critical income to support the family. A PCP explained how it affected Latino families negatively:

> *I have a large patient population that works in the meat packing plant industry, and they were really scared of COVID infection due to an outbreak they experienced.*

4.2. Disrupted Work

While this interruption gave some the opportunity to care for their loved ones, many caregivers' inability to work had a significant impact on the family's financial situation. This impact also generated family conflicts, such as the following from a caregiver who quit her job abroad to move back with her care recipient:

> *I ended up fighting with my sister, who, like me, was also living abroad because I ended up moving with Mom. I told her, who will care for Mom…our Mom?*

4.3. Increased Work Burden

Some Latino caregivers were considered essential workers during the COVID-19 pandemic, which increased their work burden and risk of infection. A PCP mentioned the following with respect to their Latino patients, including relatives of people with ADRD:

> *Many of them work in restaurants and similar jobs that have been affected by it. In construction, people have had no issues, but they have a lot more work.*

Theme 5. The Pandemic Impacted Health Care and Community Care Systems

The pandemic impacted multiple systems that support care recipients and caregivers, including health care and community care systems.

5.1. Health Care Systems

The pandemic impacted the availability and quality of services provided to Latino families with ADRD in primary, neurology, and long-term care. Services included educating Latino caregivers on ADRD, assessing for dehydration, creating rapport, and the treatment and assessment of ADRD and other conditions. PCPs had to reduce physical contact with care recipients, which reduced their chance to convey warmth to their patients. An example of how the pandemic affected services includes a PCP who said the following despite being able to see patients again after the lockdown:

> *It's not the same, you no longer have the closeness or attachment with the patient, I can no longer touch their shoulders or give them a hand because now we have to keep our distance.*

A caregiver complained about the poor prioritization of his primary care clinic with respect to obtaining the COVID-19 vaccine for people with ADRD:

> *I got an e-mail…from the hospital saying that if I wanted the COVID shot to call and make an appointment, and…I said what about my wife…she needs it more than I do.*

The novelty of COVID-19 also increased confusion between COVID-19 and ADRD symptoms in primary care and neurology clinics. A caregiver mentioned the following:

> *Both the PCP and the neurologist saw the issue as part of the ADRD, not as part of the virus…But I'm sure it was the virus because she could swallow again after two or three weeks, and she was doing better than before.*

PCPs mentioned their patients with ADRD acquiring COVID-19. As an extreme case, 1 PCP who serves institutionalized patients, which include several Latino patients, reported the following:

> *One of my nursing homes right now, I think we are up to 22 positives today out of 40 patients.*

5.2. Community Care Systems

The access and quality of services provided by community care systems to Latino families with ADRD were also impacted by

the pandemic. These included services provided by senior centers, home assistants, caregiver support groups, and home-delivered meals. Two caregivers described the following:

> *The senior center opened two weeks ago. But I told them I'd have her at home. There's no need to expose her to anything…Four hours are not going to make a difference and it's ultimately going to give me more work. And she would be exposed. I don't trust that the protocols are ideal for her to go.*

> *The reason why I'm not going to support groups now is initially because of COVID. They had stopped and I had not heard that they are doing anything with the group again.*

Speaking about their Spanish-speaking home assistants, a caregiver reported the following:

> *With COVID, this has been terrible, because people don't want to work. Mom had two that got COVID.*

Some Latino families with ADRD we interviewed tried to use home-delivered meals for the first time during the pandemic to reduce risk of infection. However, home-delivered meals services were not culturally prepared to serve Latino families, as this caregiver explained:

> *We had Meals on Wheels, but we just could not have those meals, they were terrible. They were terrible for a Latino family.*

Theme 6. Health Care and Community Care Systems Took Measures to Reduce the Impact of the Pandemic

Families and PCPs highlighted the impact Latino families with ADRD experienced and the reduced access and quality of health care and community care services. In response, these services adapted to the COVID-19 pandemic context by leveraging home health and remote communication.

6.1. Home Health

Most senior centers were closed during the pandemic, but some reopened shortly after the lockdown was lifted. Clinics and other institutions leveraged home care professionals to provide care for their patients. These included private sources but also all-inclusive care programs for older adults and Catholic charities. Regarding the all-inclusive care programs for older adults, a caregiver and a PCP mentioned the following, respectively:

> *The home care services come from an all-inclusive care program for the elderly. I think since August from last year they've been coming every day from 8 AM to 12 PM, which is when I have more work online meetings.*

> *Now if they need to be seen we send people to the home and we can do that. If somebody needs to be seen we will send a provider, a nurse, whoever else needs to see them.*

6.2. Remote Communication

Participants and PCPs stressed their increased use of remote communication during the pandemic. Purposes of remote communication included conducting telehealth visits among clinicians, caregivers, and care recipients; using interpreters; and ordering medications. The communications methods used included phone and video calls, as well as emails and patient portals. A PCP exemplified this shift as follows:

> *That's the way I've communicated with patients during the pandemic…over the phone or virtually and ordering their medications directly electronically to the pharmacy or the lab.*

Theme 7. Latino Families Experienced Barriers to Remote Communication During the Pandemic

As mentioned earlier, remote communication became paramount during the COVID-19 pandemic for health care and community care services. Remote communication also became an important tool for families. However, Latino families with ADRD experienced barriers in communication related to their low trust, insufficient skills, diminished abilities, and inadequate support.

7.1. Low Trust

Participants identified trust of the remote communication source as a barrier, specifically among the Latino families they serve. A PCP said the following:

> *I have seen specifically in the Latino population they are not quite as engaged with some of the technology and telehealth, I think they are a little bit worried about how to use it or 'is my information safe?'.*

7.2. Insufficient Skills and Diminished Abilities

Another barrier to remote communication was the low technology savviness of Latino individuals and the diminished abilities of care recipients due to their ADRD. These included cognitive or physical impairment. The relative of a care recipient said the following:

> *It is really hard to Facetime with him unless there is somebody right there with him because the camera keeps going to the ceiling or he does not want to talk to me, so having a Facetime with him is sometimes difficult.*

7.3. Inadequate Support

Families highlighted the need for support to communicate remotely, which was hindered by technological difficulties (eg, patient portals not being in Spanish and translators being harder to understand over the phone vs in person), low access to devices, and a lack of family involvement. A family caregiver who limited her visits to reduce the risk of infecting her care recipient mentioned the following:

> *I try to keep the peace with my brother, so I gave Mom an iPad, but she does not know how to use it and she needs help using it and my brother who lives with her is reluctant to facilitate the Facetime calls.*

Theme 8. Caregiver Social Support Was Critical for Reducing Social Isolation and Its Sequalae

Caregivers were crucial in protecting care recipients' health and maintaining their quality of life during the pandemic. In fact, caregiver support was the most frequently mentioned type of

social support. Caregivers largely supported care recipients in protecting against COVID-19 infection and reducing the psychosocial, health care, and community care impact of the COVID-19 pandemic.

8.1. Protecting Against COVID-19 Infection

Caregivers made sure that they and their care recipients would wear face masks and shields, wash their hands often, keep a safe distance from others (or themselves if the caregiver was exposed to others), and get tested for COVID-19. Caregivers took over some tasks they typically did not do, to increase the care recipient's safety and physical health, including taking care of medicines, monitoring nutrition, planning activities, and going shopping alone. Caregivers were proactive in acquiring vaccinations as soon as they were available for themselves and their loved ones, despite systemic barriers, which allowed them to provide care for their care recipient rather than avoiding contact. Two caregivers described how they advocated for their loved ones to get vaccinated as soon as possible:

> *I feel like if I had not been persistent and kept calling, she probably would still not have her vaccines.*
>
> *I do not understand how such an old person has not gotten a call from her primary care clinic…So, I asked a nun at a catholic hospital at a different state if she could give permission for my wife to go, and she did, so my wife got vaccinated in that other state.*

8.2. Reducing Psychosocial Impact

Social support was critical for reducing loneliness and its consequences. Caregivers sought informal sources of support, including family and friends, and formal sources, such as clinical resources and respite care. Many adult child caregivers either moved in with the care recipient or moved the care recipient to live with them to provide daily care. There were only few cases where a caregiver reported not providing family support to the care recipient. Two caregivers' quotes exemplified the importance of their social support to reduce the care recipients' loneliness and the mental health impact of the COVID-19 pandemic:

> *I am the only one that visited her and would take her anywhere. No one has visited ever since I moved back. My sister might have visited once, maybe.*
>
> *When I am not involved in calling my mother often, she is just very anxious, so I think calling multiple times a day everyday helps.*

8.3. Reducing Health Care and Community Care System Impact

Besides advocating to get their loved ones vaccinated, family caregivers had to solve additional problems related to the health care and community care systems. These problems included assisting with remote communications with their providers, requesting that their loved one's cognition is assessed, or requesting that they be allowed to accompany the care recipient into health care visits at a time when these were forbidden due to the pandemic. Two caregivers said the following:

> *The PCP told me those were side effects of COVID and that little by little her memory would come back…Every three months, I'd go to her and insist, until a neurologist saw her and diagnosed her with ADRD.*
>
> *They wouldn't let me come in with Mom, but I explained that she would not be able to understand anything due to her memory, and after some discussion, they let me in too…I convinced them despite my poor English proficiency.*

Discussion

Principal Findings

This study aimed to gain an in-depth understanding of the impact of the COVID-19 pandemic among Latino families with ADRD, from the perspectives of family caregivers and PCPs. To achieve this goal, we interviewed 21 family caregivers of Latino individuals with ADRD and 23 PCPs across the United States. These participants were diverse with respect to their region, primary language, and other characteristics. We found that the COVID-19 pandemic has impacted Latino families with ADRD at multiple levels, ranging from physical to health care and community care levels. Latino family caregivers' actions were key to addressing these new challenges.

Comparison With Prior Work

Some impacts of the COVID-19 pandemic on Latino families with ADRD resembled impacts reported in previous research for the general population. For example, accelerated cognitive deterioration was also reported in a mixed methods study in the United Kingdom where 184 caregivers and 24 people with ADRD were interviewed [22]. The lack of access to engaging activities and human contact were also reported in the UK study, which were found to be potential drivers of the accelerated deterioration. Other similar findings from this study include the impact on care recipients' and caregivers' mental health, the barriers with the technology needed to replace in-person contacts, and the importance of caregivers to bring about resilience despite the adversities. Similar impacts have also been reported among family caregivers of people with ADRD in India and Italy [23,24]. Despite these similarities, our findings highlight the need to pay attention to pandemic-related stress that add up with many other life stresses found frequently in minoritized populations [25], as well as additional factors that compound the already existing digital divide's impact on remote communication among Latino individuals: low trust in these technologies and decline in technological abilities among people with ADRD [13,23,24].

Some impacts have been reported less frequently or not reported at all previously. To our knowledge, no study has reported the impact of the COVID-19 pandemic on nutrition among people with ADRD and their families. This impact is in line with a systematic review of 28 studies across the world, where most of them identified changes in food intake toward the adoption of unhealthy eating among people of all ages [26]. Our findings shed light on potential solutions for the Latino community, as those impacted tended to be male caregivers with low income

and families for which the meal delivery services did not provide culturally appropriate meals.

We also found that the COVID-19 pandemic impacted caregivers' work. This finding aligns with a study from India where many caregivers had to start working from home or faced disruptions that affected their finances [24]. In our Latino sample, the pandemic did not just disrupt work for some but also increased the work burden of others. Latino individuals may have been especially impacted by the work consequences of the pandemic, as they are overrepresented in the frontline workforce, tend to live in multigenerational homes, and have among the lowest average incomes in the United States [27-29].

Our findings align with 2 commonly reported factors among Latino individuals that have been shown to impact health behaviors, including health care use and assertiveness [30-32]. First, our interviews revealed that the news about Latino individuals' disproportionate risk of COVID-19 infection and death [9] may have increased the fear among Latino families and contributed to their low use of sometimes critically needed resources. These findings align with "fatalism," a belief that events are beyond one's control [31]. Second, we also found that to reduce the risk of infection, PCPs reduced physical contact with their patients, which they saw as a barrier to care. This finding suggests an impact on "personalism," or the importance of establishing relationships through warm interactions [31].

Limitations

This study has some limitations. Remote recruitment and interviews increased the representation of participants in rural areas and other states. However, video calls and phone calls led to some communication issues, which, in some cases, reduced the amount of information we could collect and affected the quality of the audio. The inability to conduct in-person recruitment and interviews may have excluded the most underserved individuals, who could have been contacted via health fairs before the pandemic started. We did not interview individuals with ADRD, which did not allow a full triangulation between them, their caregivers, and PCPs. Although Latino caregivers tend to be women [33], they were overrepresented in our study, likely also due to women's higher likelihood to participate in health-related research [34,35]. The sample size was relatively small and not probabilistic, which reduces the generalizability of the findings. We did not return the transcripts to participants for comment or correction. As with most studies, individuals who participated in the study were motivated to participate. We do not know how much their discourse compares to those who decided not to participate.

Implications and Future Directions

This study has implications for public health. Given the efficacy of existing COVID-19 vaccines [36], ensuring access to ongoing boosters among Latino individuals with ADRD and their families will be needed. To do so, it will continue to be necessary to hold events at flexible times and days; provide care at convenient venues; and improve the communication with Latino individuals by using a wide range of communication modalities (eg, calling, texting, and patient portals) in a way that is trustworthy, easy to use for people with cognitive impairment, and linguistically and culturally appropriate. The common stress and depression related to the fear of COVID-19 infection, uncertainty, and confinement among Latino families with ADRD highlight the need improve access to mental health services in general and specifically during pandemics. An example of a potentially inexpensive intervention is layperson-delivered, empathy-oriented telephone call programs, which have been shown to reduce loneliness, depression, and anxiety [37]. These services can be provided by governments or charity organizations and offered via primary care clinics, health departments, and entities that identify those who are potentially in need. Other useful services may include cognitive and physical engagement activities to reduce confinement-related deterioration. Since stresses are cumulative and Latino individuals tend to have higher levels of socioeconomically driven chronic stress [25], building stronger and more accessible social and health care welfare systems will also be key.

The impact of the pandemic on physical and cognitive decline, potential sources of support, and the crucial importance of families in the care of Latino individuals with ADRD highlight the need to provide them with financial support for their services, respite when possible, and accessible caregiving support training. Home care services have been essential in caring for some Latino individuals with ADRD and providing respite to their caregivers. However, these professions tend to be poorly paid, and their services are hard to access. Home care workers' salary could be adjusted to the value of their service and covered by health insurance companies or government programs.

Our findings can inform future studies. Our findings regarding the physical and cognitive deterioration caused by the pandemic and the importance of family support may help inform future studies on the differential impact on ethnic groups in community and institutional settings and on the buffering effect of social support and health care and community care services. For example, nursing homes that have a higher proportion of ethnically minoritized populations had higher rates of COVID-19 infection and death [38]. Studies could explore whether this disproportionate impact also applies to cognitive and functional decline. Our data suggest that Latino families where men were the primary caregiver were at risk of poor nutrition and that home meal delivery services not offering culturally appropriate meals was a barrier to improving nutrition quality. Future studies can develop remote interventions to provide education on how to cook healthy, culturally appropriate meals for Latino family caregivers and explore the cost-effectiveness of offering Latin American healthy meals in communities where Latino individuals are present. Our findings also suggested that the awareness that Latino individuals had a higher infection and death risk from COVID-19 might have reduced their use of health care services. Future studies should develop culturally appropriate messaging in response to disparities where the fear and related factors facilitate health behaviors instead of reducing them.

Conclusion

In this study, we have found that the COVID-19 pandemic has impacted Latino families with ADRD beyond infection and physical symptoms, and family caregivers have been crucial to maintaining care recipients' health and quality of life. The experiences of Latino families with ADRD during the pandemic resembled those of the general population. These experiences included a lack of access to engaging activities and human contact, issues with the technology needed to replace in-person contacts, accelerated cognitive deterioration, and an impact on caregivers' and care recipients' mental health. However, this pandemic has revealed many of the barriers that Latino families with ADRD face, and in most cases, this pandemic has exacerbated previous barriers. These barriers included nutrition being affected by inefficiencies of the mailing system or financial insecurity; finances being affected by the pandemics' impact on jobs typically held by Latino caregivers; and fatalism and personalism interacting with Latino individuals' disproportionate risk to COVID-19 infection and death, which reduced health care use. These findings suggest the need for more equitable access, culturally appropriate and trustworthy content and delivery of health care and community services, as well as stronger financial and social supports for family caregivers.

Acknowledgments

This work was supported by grants K01 MD014177 and P30 AG072973 from the National Institutes of Health. The authors are thankful to Ms Licet Valois; R.A.I.N. Total Care Inc; Children's Mercy; and other people and organizations that assisted with the recruitment of participants. The authors are also grateful to all the participants from this study, especially given the difficult circumstances due to the COVID-19 pandemic.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guides.
[DOCX File, 18 KB - xmed_v5i1e42211_app1.docx]

References

1. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. Alzheimers Dement 2021 Mar 23;17(3):327-406. [doi: 10.1002/alz.12328] [Medline: 33756057]
2. Lin PJ, Emerson J, Faul JD, et al. Racial and ethnic differences in knowledge about one's dementia status. J Am Geriatr Soc 2020 Aug;68(8):1763-1770. [doi: 10.1111/jgs.16442] [Medline: 32282058]
3. Barker WW, Luis C, Harwood D, et al. The effect of a memory screening program on the early diagnosis of Alzheimer disease. Alzheimer Dis Assoc Disord 2005;19(1):1-7. [doi: 10.1097/01.wad.0000155380.63231.68] [Medline: 15764864]
4. Mehta KM, Yin M, Resendez C, Yaffe K. Ethnic differences in acetylcholinesterase inhibitor use for Alzheimer disease. Neurology 2005 Jul 12;65(1):159-162. [doi: 10.1212/01.wnl.0000167545.38161.48] [Medline: 16009909]
5. Scharlach AE, Giunta N, Chow JCC, Lehning A. Racial and ethnic variations in caregiver service use. J Aging Health 2008 Apr;20(3):326-346. [doi: 10.1177/0898264308315426] [Medline: 18332187]
6. Faison WE, Schultz SK, Aerssens J, et al. Potential ethnic modifiers in the assessment and treatment of Alzheimer's disease: challenges for the future. Int Psychogeriatr 2007 Jun;19(3):539-558. [doi: 10.1017/S104161020700511X] [Medline: 17451614]
7. Query system. National Alzheimer's Coordinating Center. URL: https://www.alz.washington.edu/WEB/naccquery.html [accessed 2018-12-21]
8. Connell CM, Scott Roberts J, McLaughlin SJ. Public opinion about Alzheimer disease among Blacks, Hispanics, and Whites: results from a national survey. Alzheimer Dis Assoc Disord 2007;21(3):232-240. [doi: 10.1097/WAD.0b013e3181461740] [Medline: 17804956]
9. Demographic trends of COVID-19 deaths in the US reported to NVSS. Centers for Disease Control and Prevention. 2022. URL: https://covid.cdc.gov/covid-data-tracker/#demographics [accessed 2022-8-12]
10. Painter EM, Ussery EN, Patel A, et al. Demographic characteristics of persons vaccinated during the first month of the COVID-19 vaccination program—United States, December 14, 2020-January 14, 2021. MMWR Morb Mortal Wkly Rep 2021 Feb 5;70(5):174-177. [doi: 10.15585/mmwr.mm7005e1] [Medline: 33539333]
11. de Snyder VNS, McDaniel M, Padilla AM, Parra-Medina D. Impact of COVID-19 on Latinos: a social determinants of health model and scoping review of the literature. Hisp J Behav Sci 2021 Aug;43(3):174-203. [doi: 10.1177/07399863211041214]
12. Lines LM, Wiener JM, US Department of Health and Human Services, Assistant Secretary for Planning and Evaluation, Office of Disability, Aging and Long-Term Care Policy. Racial and ethnic disparities in Alzheimer's disease: a literature

review. Office of the Assistant Secretary for Planning and Evaluation. 2014 Feb. URL: https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//138596/RacEthDis.pdf [accessed 2024-02-15]
13. Portacolone E, Chodos A, Halpern J, et al. The effects of the COVID-19 pandemic on the lived experience of diverse older adults living alone with cognitive impairment. Gerontologist 2021 Feb 23;61(2):251-261. [doi: 10.1093/geront/gnaa201] [Medline: 33404634]
14. Gallagher-Thompson D, Solano N, Coon D, Areán P. Recruitment and retention of Latino dementia family caregivers in intervention research: issues to face, lessons to learn. Gerontologist 2003 Feb;43(1):45-51. [doi: 10.1093/geront/43.1.45] [Medline: 12604745]
15. Drabo EF, Barthold D, Joyce G, Ferido P, Chang Chui H, Zissimopoulos J. Longitudinal analysis of dementia diagnosis and specialty care among racially diverse Medicare beneficiaries. Alzheimers Dement 2019 Nov;15(11):1402-1411. [doi: 10.1016/j.jalz.2019.07.005] [Medline: 31494079]
16. Schwandt TA. Dictionary of Qualitative Inquiry, 2nd edition: SAGE Publishing; 2001.
17. Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The National Institute on Minority Health and Health Disparities research framework. Am J Public Health 2019 Jan;109(S1):S16-S20. [doi: 10.2105/AJPH.2018.304883] [Medline: 30699025]
18. Pollock T. The difference between structured, unstructured & semi-structured interviews. Oliver Parks. URL: https://www.oliverparks.com/blog-news/the-difference-between-structured-unstructured-amp-semi-structured-interviews [accessed 2024-02-15]
19. Miles MB, Huberman AM. Qualitative Data Analysis: A Source Book of New Methods, 1st edition: Sage Publications; 1984.
20. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52. [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
21. Taylor SS, Treacy A. Just Dedoose it! making mixed methods data analysis manageable. Presented at: Annual Meeting of the Northern Rocky Mountain Educational Research Association; Oct 3, 2013; Jackson Hole, WY URL: https://www.academia.edu/6415766/Just_Dedoose_It_Making_Mixed_Methods_Data_Analysis_Manageable [accessed 2024-02-15]
22. Harding E, Rossi-Harries S, Gerritzen EV, et al. "I felt like I had been put on the shelf and forgotten about" - lasting lessons about the impact of COVID-19 on people affected by rarer dementias. BMC Geriatr 2023 Jun 27;23(1):392. [doi: 10.1186/s12877-023-03992-1] [Medline: 37370011]
23. Cipolletta S, Morandini B, Tomaino SCM. Caring for a person with dementia during the COVID-19 pandemic: a qualitative study with family care-givers. Ageing Soc 2023 Mar;43(3):535-555. [doi: 10.1017/S0144686X21000696]
24. Vaitheswaran S, Lakshminarayanan M, Ramanujam V, Sargunan S, Venkatesan S. Experiences and needs of caregivers of persons with dementia in India during the COVID-19 pandemic-a qualitative study. Am J Geriatr Psychiatry 2020 Nov;28(11):1185-1194. [doi: 10.1016/j.jagp.2020.06.026] [Medline: 32736918]
25. Brown LL, Mitchell UA, Ailshire JA. Disentangling the stress process: race/ethnic differences in the exposure and appraisal of chronic stressors among older adults. J Gerontol B Psychol Sci Soc Sci 2020 Feb 14;75(3):650-660. [doi: 10.1093/geronb/gby072] [Medline: 29878196]
26. Penna PM, Oliveira NMC, Castro LCV, Hermsdorff HHM. Consequences of quarantine during the COVID-19 pandemic on food intake and body weight: a systematic review. Research Square. Preprint posted online on Oct 28, 2021. [doi: 10.21203/rs.3.rs-953160/v1]
27. Goldman N, Pebley AR, Lee K, Andrasfay T, Pratt B. Racial and ethnic differentials in COVID-19-related job exposures by occupational standing in the US. PLoS One 2021 Sep 1;16(9):e0256085. [doi: 10.1371/journal.pone.0256085] [Medline: 34469440]
28. Semega J, Kollar M, Shrider EA, Creamer JF. Income and poverty in the United States: 2019. US Census Bureau. 2020 Sep. URL: https://www.census.gov/content/dam/Census/library/publications/2020/demo/p60-270.pdf [accessed 2024-02-15]
29. Cohn D, Passel JS. A record 64 million Americans live in multigenerational households. Pew Research Center. 2018 Apr 5. URL: https://www.pewresearch.org/fact-tank/2018/04/05/a-record-64-million-americans-live-in-multigenerational-households/ [accessed 2024-02-15]
30. de Santis JP, Gattamorta KA, Valdes B, Sanchez M, Provencio-Vasquez E. The relationship of Hispanic cultural factors and sexual behaviors of Hispanic men who have sex with men. Sex Cult 2019 Mar;23(1):292-309. [doi: 10.1007/s12119-018-9557-8] [Medline: 37090487]
31. Bermúdez JM, Kirkpatrick DR, Hecker L, Torres-Robles C. Describing Latinos families and their help-seeking attitudes: challenging the family therapy literature. Contemp Fam Ther 2010 Jun;32(2):155-172. [doi: 10.1007/s10591-009-9110-x]
32. de Los Monteros KE, Gallo LC. The relevance of fatalism in the study of Latinas' cancer screening behavior: a systematic review of the literature. Int J Behav Med 2011 Dec;18(4):310-318. [doi: 10.1007/s12529-010-9119-4] [Medline: 20953916]
33. Evercare, National Alliance for Caregiving. Evercare® study of Hispanic caregiving in the U.S. National Alliance for Caregiving. 2008 Nov. URL: https://www.caregiving.org/wp-content/uploads/2020/05/Hispanic_Caregiver_Study_web_ENG_FINAL_11_04_08.pdf [accessed 2024-02-15]
34. Perales-Puchalt J, Shaw A, McGee JL, et al. Preliminary efficacy of a recruitment educational strategy on Alzheimer's disease knowledge, research participation attitudes, and enrollment among Hispanics. Hisp Health Care Int 2020 Sep;18(3):144-149. [doi: 10.1177/1540415319893238] [Medline: 31840539]

35. Perales-Puchalt J, Barton K, Ptomey L, et al. Effectiveness of “reducing disability in Alzheimer's disease” among dyads with moderate dementia. J Appl Gerontol 2021 Oct;40(10):1163-1171. [doi: 10.1177/0733464820934683] [Medline: 32608300]
36. Lopez Bernal J, Andrews N, Gower C, et al. Effectiveness of COVID-19 vaccines against the B.1.617.2 (Delta) variant. N Engl J Med 2021 Aug 12;385(7):585-594. [doi: 10.1056/NEJMoa2108891] [Medline: 34289274]
37. Kahlon MK, Aksan N, Aubrey R, et al. Effect of layperson-delivered, empathy-focused program of telephone calls on loneliness, depression, and anxiety among adults during the COVID-19 pandemic: a randomized clinical trial. JAMA Psychiatry 2021 Jun 1;78(6):616-622. [doi: 10.1001/jamapsychiatry.2021.0113] [Medline: 33620417]
38. Gorges RJ, Konetzka RT. Factors associated with racial differences in deaths among nursing home residents with COVID-19 infection in the US. JAMA Netw Open 2021 Feb 1;4(2):e2037431. [doi: 10.1001/jamanetworkopen.2020.37431] [Medline: 33566110]

Abbreviations

ADRD: Alzheimer disease and related dementias
PCP: primary care provider

Edited by E Meinert, G Eysenbach; submitted 26.08.22; peer-reviewed by A Marin, C Peterson, Anonymous; revised version received 11.12.23; accepted 13.01.24; published 08.03.24.

Please cite as:
Perales-Puchalt J, Peltzer J, Fracachan-Cabrera M, Perez GA, Ramírez M, Greiner KA, Burns JM
Impact of the COVID-19 Pandemic on Latino Families With Alzheimer Disease and Related Dementias: Qualitative Interviews With Family Caregivers and Primary Care Providers
JMIRx Med 2024;5:e42211
URL: https://xmed.jmir.org/2024/1/e42211
doi:10.2196/42211

Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation

Izzy Kohler[1], DPT; Michael V Perrotta[1], BSc; Tiago Ferreira[1], MS; David M Eagleman[1,2], PhD

[1]

[2]

Corresponding Author:
David M Eagleman, PhD

Related Articles:

Companion article: http://preprints.jmir.org/preprint/49969

Companion article: https://www.medrxiv.org/content/10.1101/2023.06.01.23290351v1

Companion article: https://med.jmirx.org/2024/1/e55554

Companion article: https://med.jmirx.org/2024/1/e55727

Companion article: https://med.jmirx.org/2024/1/e55728

Companion article: https://med.jmirx.org/2024/1/e55510

Abstract

Background: High-frequency hearing loss is one of the most common problems in the aging population and with those who have a history of exposure to loud noises. This type of hearing loss can be frustrating and disabling, making it difficult to understand speech communication and interact effectively with the world.

Objective: This study aimed to examine the impact of spatially unique haptic vibrations representing high-frequency phonemes on the self-perceived ability to understand conversations in everyday situations.

Methods: To address high-frequency hearing loss, a multi-motor wristband was developed that uses machine learning to listen for specific high-frequency phonemes. The wristband vibrates in spatially unique locations to represent which phoneme was present in real time. A total of 16 participants with high-frequency hearing loss were recruited and asked to wear the wristband for 6 weeks. The degree of disability associated with hearing loss was measured weekly using the Abbreviated Profile of Hearing Aid Benefit (APHAB).

Results: By the end of the 6-week study, the average APHAB benefit score across all participants reached 12.39 points, from a baseline of 40.32 to a final score of 27.93 (SD 13.11; N=16; *P*=.002, 2-tailed dependent *t* test). Those without hearing aids showed a 10.78-point larger improvement in average APHAB benefit score at 6 weeks than those with hearing aids (t_{14}=2.14; *P*=.10, 2-tailed independent *t* test). The average benefit score across all participants for ease of communication was 15.44 (SD 13.88; N=16; *P*<.001, 2-tailed dependent *t* test). The average benefit score across all participants for background noise was 10.88 (SD 17.54; N=16; *P*=.03, 2-tailed dependent *t* test). The average benefit score across all participants for reverberation was 10.84 (SD 16.95; N=16; *P*=.02, 2-tailed dependent *t* test).

Conclusions: These findings show that vibrotactile sensory substitution delivered by a wristband that produces spatially distinguishable vibrations in correspondence with high-frequency phonemes helps individuals with high-frequency hearing loss improve their perceived understanding of verbal communication. Vibrotactile feedback provides benefits whether or not a person wears hearing aids, albeit in slightly different ways. Finally, individuals with the greatest perceived difficulty understanding speech experienced the greatest amount of perceived benefit from vibrotactile feedback.

(JMIRx Med 2024;5:e49969) doi:10.2196/49969

KEYWORDS

audiology; hearing; high-frequency; wristband; develop; development; wearable; wearables; machine learning; phoneme; phonemes; hear; vibrotactile; vibration; vibrations; sound; sounds; hearing loss; loud noise; loud noises; noise pollution; hearing aids; hearing aid

Introduction

Hearing loss affects 466 million people worldwide [1]. High-frequency hearing loss is one of the most common types of hearing loss and renders high-pitched sounds, such as the voices of women and children, more difficult to hear [2,3]. It can affect people of any age but is more common among older adults and people who have been repeatedly exposed to loud noises [4-6]. This type of hearing loss can be frustrating and disabling, making it difficult to understand speech communication and interact effectively with the world, leading to a decline in quality of life and isolation [6,7].

Individuals with high-frequency hearing loss struggle to hear consonants with higher-frequency sound components, such as *s*, *t*, and *f*. As a result of the hearing loss, speech is reported as sounding muffled, most noticeably in noisy environments. Commonly, people with high-frequency hearing loss will report that they can hear but cannot understand [8]. It is often noticed when a person has trouble understanding women's and children's voices and detecting other sounds such as the ringing of a cell phone or the chirping of birds. Assistive hearing technologies such as hearing aids and cochlear implants can offer some assistance with understanding speech communication, but they have limitations. One of the most commonly reported disappointments among users of hearing aids and cochlear implants is that they still cannot understand speech, especially in complex environments [9,10].

To address the speech understanding limitations associated with high-frequency hearing loss, we have developed a vibrotactile sensory substitution solution in the form of a wristband [11-13]. This device delivers spatially unique vibrations to the wrist in correspondence with target phonemes that are commonly difficult for individuals with presbycusis to detect. The wristband receives sound from the environment through an onboard microphone and uses a machine learning algorithm to filter background noise (BN) and extract target phonemes from speech. Each phoneme signal is mapped to its own unique linear resonant actuator (LRA) in the strap of the wristband where it is felt as a vibration on the skin. There are four LRAs embedded within the wristband strap, giving each target phoneme a unique spatial location on the wrist. Parts of speech that are audible to the user are unconsciously integrated with the spatially unique vibratory signals representing the inaudible portions of speech. The user is then able to understand a complete and meaningful message through the integration of the complementary sensory inputs [11-13].

Our prior work in this area demonstrated that when two words are algorithmically translated into spatiotemporal patterns of vibration on the skin of the wrist, they are distinguishable to individuals who are hard of hearing or deaf up to 83% of the time for two words that are similar and up to 100% of the time for two words that are not similar [12,14]. Further studies showed that sound-to-touch sensory substitution devices may help people with hearing impairments, allowing them to access sensory information that is otherwise inaccessible. Weisenberger and Russell [15] used single-channel vibrotactile aids designed to translate acoustic stimuli into representative vibration patterns on the wrist to improve performance on environmental sound identification tests from 55% to 95% correct and improve performance on single word identification testing from 60% to 90%.

In this study, we aimed to demonstrate that a simple wearable sensory substitution device that transforms speech sounds into haptic vibrations on the wrist can help individuals with high-frequency hearing loss perceive a greater ability to understand speech communication throughout their normal daily routine. With further development and refinement, this technology has the potential to improve the quality and productivity of their daily interactions, enable them to enjoy audio-based entertainment such as movies and podcasts, help them understand conversations in complicated acoustic environments, and fill the residual gaps of impairment left by their hearing aids.

Methods

Participants

Participants were recruited via web-based advertising for a paid study related to hearing loss. Eligibility required (1) an age between 18 and 80 years, (2) having access to a mobile device (iOS or Android) and a computer, (3) English as a primary spoken language, and (4) meeting the following criteria for high-frequency hearing loss: a pure-tone audiogram (either from an audiologist in the past 24 mo or from 2 audiogram mobile apps, Mimi and Hearing & Ear Age Test) must show at least 55 dB of hearing loss at 4 kHz averaged across both ears (with neither ears' 4-kHz threshold being less than 40 dB of hearing loss) and no more than 35 dB of hearing loss averaged across both ears and across 500-Hz and 1000-Hz tones. These specifications were chosen to capture individuals with hearing loss profiles in alignment with high-frequency hearing loss. Candidates who did not have an audiogram from an audiologist were required to provide audiograms from both audiogram mobile apps, which have been demonstrated as comparable to in-clinic testing [16].

A total of 16 eligible participants completed the study: 10 male participants, 5 female participants, and 1 nonbinary participant. The average age was 68.8 (SD 11.6) years. The type and severity of hearing loss were determined from pure-tone audiograms. A total of 9 participants provided audiograms from an audiologist and 7 provided audiograms from the two mobile apps. The average pure-tone threshold of both ears at 500 Hz and 1000 Hz was 30 (SD 13) dB and the average pure-tone threshold of both ears at 4000 Hz was 63 (SD 9) dB of hearing loss (Figure 1). Demographic data for the participants is shown in Table 1.

Figure 1. Average pure-tone audiogram of both ears. The thin lines represent each participant; the thick line represents the group average. HL: hearing loss.

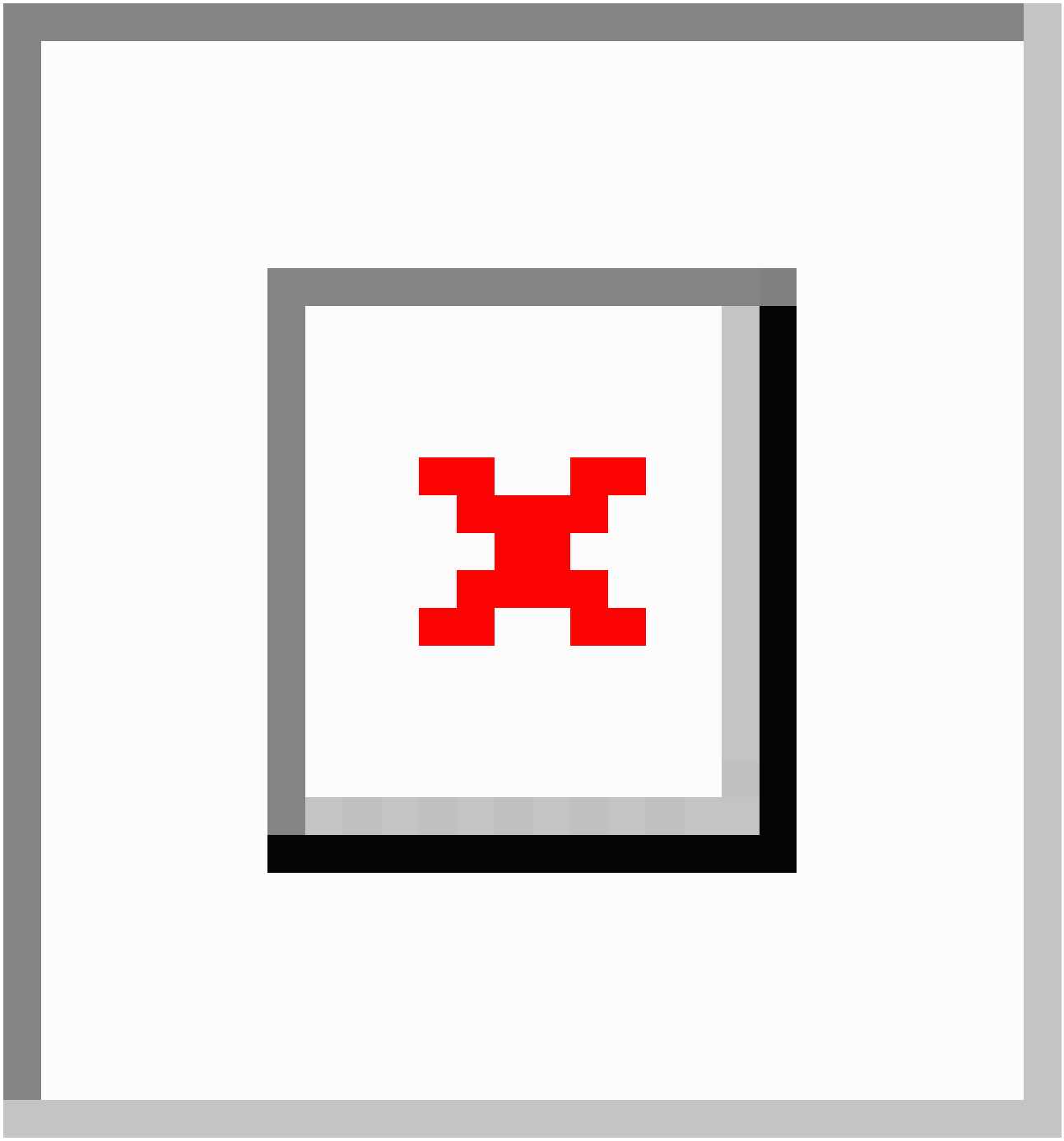


Table . Demographic data.

	Age (y)	Sex	Hearing aids	Years with hearing loss	Audiogram source[a]	Hearing loss (dB)[b]											
						250		500		1000		2000		4000		8000	
						R[c]	L[d]	R	L	R	L	R	L	R	L	R	L
B1	64	Male	No	4	Audiologist	10	15	10	15	15	20	15	20	50	55	55	80
B2	75	Male	Yes	15	Audiologist	20	35	15	25	15	15	30	50	80	75	80	85
B3	72	Nonbinary	Yes	22	Audiologist	10	25	25	40	45	45	50	55	65	70	100	100
B4	69	Female	No	35	Audiologist	35	35	45	40	45	45	55	60	55	65	55	55
B5	74	Female	No	15	Mobile app	30	28	33	28	20	20	40	35	55	53	70	70
B6	78	Female	Yes	10	Mobile app	20	20	28	28	48	48	63	73	78	80	70	70
B7	27	Male	Yes	10	Audiologist	10	10	5	5	0	0	5	5	80	80	80	80
B8	73	Male	No	3	Mobile app	33	30	38	35	40	45	45	45	58	60	70	65
B9	68	Male	Yes	15	Mobile app	33	25	40	33	45	38	48	43	58	58	70	65
B10	67	Female	Yes	10	Mobile app	35	33	43	35	50	48	48	48	55	53	55	70
B11	76	Male	Yes	25	Mobile app	33	30	18	23	13	38	58	60	68	68	70	70
B12	66	Female	No	5	Audiologist	25	25	35	35	40	40	60	60	60	60	80	70
B13	79	Male	Yes	15	Audiologist	40	35	35	35	40	35	65	50	65	60	65	60
B14	67	Male	Yes	10	Audiologist	5	5	5	10	10	10	25	30	55	60	75	75
B15	74	Male	No	5	Mobile app	28	20	43	30	33	25	45	45	65	68	65	70
B16	71	Male	No	20	Audiologist	40	35	40	35	40	40	45	50	50	60	60	60

[a]Audiogram source indicates where the audiogram originated from. Audiologist indicates the audiogram was measured by an audiologist, and mobile app indicates the participant provided two audiograms measured by the Mimi and Hearing & Ear Age Test mobile apps.

[b]Decibels of hearing loss at 7 pure tones in the left and right ears. Hearing loss values are measured without cochlear implants or hearing aids. Note, 90 dB of hearing loss is the most the test can detect.

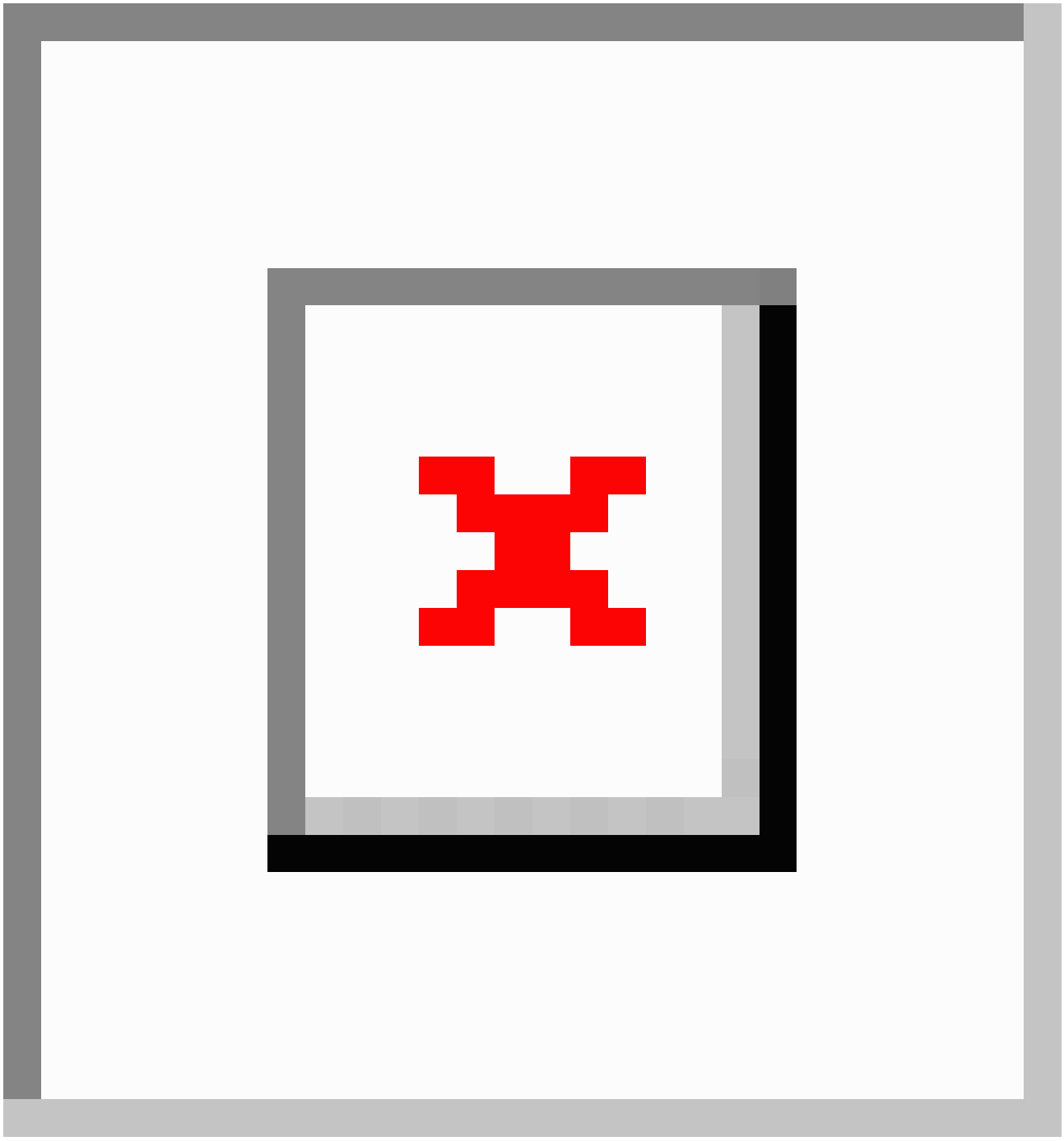
[c]R: right.

[d]L: left.

Device

Participants wore a haptic wristband (Figure 2) that vibrated to indicate the occurrence of specific phonemes. The wristband contained four vibrating motors embedded in the wrist strap, a microphone, a power button, a microcontroller, and a battery.

Figure 2. The Neosensory wristband has four vibrating motors embedded in the wrist strap. The top of the wristband contains a power button and a microphone. Each phoneme is assigned to an independent motor.



The motors were LRAs that vibrated in a sine wave and were capable of rising from 0% to 50% of their maximum amplitude within 30 milliseconds. The motors vibrated at 175 Hz, the frequency at which human skin has the highest sensitivity [17]. Each motor vibrated at 1.7 GRMS (root mean squared

acceleration from gravity; 16.6 m/s^2). The motors were separated from one another at a distance of 18.2 mm and 19.2 mm for the small and large wristband sizes, respectively (center-to-center distances). Each motor pad contacted the wearer's skin on a rectangular area that measured approximately 8.2 mm by 8.5 mm.

The top of the wristband was a module that contained the power button, a microphone, and a microcontroller. The microphone captured audio and sent this data to the microcontroller. The microcontroller processed the audio data through a phoneme-detection algorithm and vibrated the motors according to the output of the algorithm. Additional microphone characteristics are provided in Multimedia Appendix 1.

Algorithm

The algorithm processed incoming audio to determine when any target phoneme was detected. If a target phoneme was detected, the corresponding motor vibrated for 80 ms.

The four target phonemes were /s/, /t/, /z/, and /k/. Each motor on the wristband was assigned to a different target phoneme. Figure 2 shows the motor assignments for each phoneme. The four phonemes were chosen based on a combination of the following three factors: (1) how difficult each phoneme is for hearing-impaired listeners to hear, (2) how frequently each phoneme occurs in spoken English, and (3) how well our algorithm can detect each phoneme. The difficulty was pooled from several studies of phoneme confusion for hearing-impaired listeners. Phatak et al [18] asked older hearing-impaired listeners to identify the consonant in a presented consonant-vowel syllable. Woods et al [19] presented the California Syllable Test, which uses consonant-vowel-consonant syllables, to older hearing-impaired listeners in both aided (with hearing aids) and unaided conditions. Sher and Owens [20] presented a four-alternative forced-choice test with consonant-vowel-consonant syllables, where either the initial or final consonant differed between choices. Synthesizing the results of these three studies, we found that the following consonants are the most difficult to hear for a listener with presbycusis: /dh/, /th/, /ng/, /v/, /b/, /hh/, /f/, /z/, /s/, and /t/. Of these, /th/ and /ng/ are present in spoken English less than 1% of the time [21]. Our algorithm performed poorly on /dh/, /b/, /f/, and /hh/.

Phoneme Detection

The phoneme detection algorithm was trained using the elastic compute cloud on Amazon Web Services. The training data consisted of a combination of pure LibriSpeech and LibriSpeech rerecorded through the onboard microphone on the wristband. LibriSpeech is a corpus of approximately 1000 hours of English speech with standard American accents sampled at 16 kHz that has been shown to produce excellent performance in speech recognition models trained with it [22]. To produce a corpus of English read speech suitable for training speech recognition systems, LibriSpeech aligns and segments audiobook read speech with the corresponding book text automatically and then filters out portions with noisy transcripts. The purpose of using rerecorded data was to tune the algorithm's parameters to speech sounds representative of those it would encounter from the wristband's microphone.

The algorithm consisted of feature extraction and inference engine components. The feature extraction module segmented an audio stream captured from the microphone into 32-millisecond frames with 16 milliseconds of overlap. Each audio frame underwent analysis to extract distinct features suitable for phoneme recognition. The features were also subject to further processing that amplified phoneme-specific information contained and ensured robustness toward continuously changing environmental conditions.

The inference engine took these feature vectors and output phoneme predictions. The core of the inference engine was a neural network model that used a real-time temporal convolutional network structure optimized for real-time speech recognition. The full latency from phoneme onset to vibration onset was 170 milliseconds. The algorithm performance is shown below in Table 2.

Table . Algorithm performance.

	Precision[a]	Recall[b]	F_1-score[c]
K	0.86	0.75	0.8
S	0.86	0.89	0.88
T	0.85	0.65	0.74
Z	0.86	0.72	0.78
Macroaverage	0.86	0.75	0.8

[a]Precision is the ability of a classification model to return only the data points in a class. It is calculated by dividing the true positives by the sum of the true positives and false positives.

[b]Recall is the ability of a classification model to identify all data points in a relevant class. It is calculated by dividing the true positives by the sum of the true positives and false negatives.

[c]F_1-scores are a single metric that combines recall and precision using the harmonic mean. It is calculated by dividing the true positives by the sum of the true positives plus half of the sum of the false positives and false negatives.

Paradigm

Participants wore the wristband every day for 6 weeks. Each day the participants were required to spend at least 1 hour watching television or listening to an audiobook, podcast, or other speech-based media while wearing the wristband and not wearing earbuds or headphones. The instructions were to choose something engaging so their attention would be directed toward understanding what was being said, while the wristband provided the assistive haptic feedback. No further guidelines were enforced for distance from the audio source or volume. The purpose of this required daily exercise was to ensure the participant was immersed in a minimum amount of active listening each day so the brain would learn to integrate the audible speech sounds with the haptic vibratory representations of the inaudible speech sounds to form a complete meaning. In addition to the required hour of practice, participants were encouraged to wear the wristband whenever engaged in conversation or active listening to speech communication.

Tasks

Abbreviated Profile of Hearing Aid Benefit

Before starting the study and at the end of each week during the study, participants completed a modified version of the Abbreviated Profile of Hearing Aid Benefit (APHAB) that did not include 6 questions related to the aversiveness subscale [11]. These questions were removed because they ask about the unpleasantness of sounds heard through a hearing aid, which does not apply to our device. The remaining 18 questions on the APHAB ask questions about one's ability to understand verbal communication in different scenarios. For example, one of the questions is "When I am in a crowded grocery store, talking with the cashier, I can follow the conversation." In the conventional questionnaire, participants answer the questions independently about their experiences while using and while not using their hearing aids. In this study, participants answered the questions independently about their experiences while using and while not using the wristband. If the participant regularly wore hearing aids, "with the wristband" referred to wearing the wristband in addition to their hearing aids, and "without the wristband" referred to wearing their hearing aids alone. The test was administered through a web-based questionnaire that captured the data onto a datasheet for analysis. The benefit score is calculated by subtracting the final aided score at the conclusion of the trial from the baseline unaided score that was measured at the beginning of the trial. Lower raw APHAB scores indicate lower levels of disability associated with hearing loss. Higher benefit scores indicate more perceived benefits from the intervention.

Final Questionnaire

On the final day of the study, participants answered a questionnaire that asked 2 questions using a Likert scale from 1 to 10: "How much did the Clarify wristband help you understand speech?" and "How likely are you to recommend the Clarify wristband to a friend or colleague?"

Ethical Considerations

The study protocol was approved by Solutions IRB (Protocol #2016/01/7), an independent institutional review board accredited by the Association for the Accreditation of Human Research Protection Programs, Inc. All participants gave written informed consent following the Declaration of Helsinki. Upon completion of the study, participants were given a US $100 Amazon gift card for their time. At the conclusion of the study, all data were deidentified to safeguard participant information.

Results

As shown in Figure 3 and Multimedia Appendix 2, after only 1 week of wearing the wristband daily, the average APHAB benefit score (*unaided – aided*) was 8.61 points, with a baseline score of 40.32 points that dropped to 31.71 points (SD 12.11; N=16; *P*=.01, 2-tailed dependent *t* test). Baseline was defined as the unaided APHAB score taken before starting to use the wristband. As a reminder, if the participant regularly used hearing aids, they were asked to answer the unaided questions based on how they felt with their hearing aids on. If the participants never used hearing aids, they were asked to answer the unaided questions based on how they felt without any hearing assistance. The average aided APHAB score continued to trend down for the remaining 5 weeks of the study. By the end of the 6-week study, the average APHAB benefit score had reached a clinically meaningful and statistically significant value of 12.39 points [23] from a baseline of 40.32 to a final score of 27.93 (SD 13.11; N=16; *P*=.002, 2-tailed dependent *t* test). Individual data is presented in Figure 4.

Figure 3. Six-week progression of the APHAB scores. Error boundary (light blue) represents SE of the mean. Week 0 score is the unaided APHAB score (before starting with the wristband); subsequent weeks show the aided APHAB score with the wristband. APHAB: Abbreviated Profile of Hearing Aids Benefit.

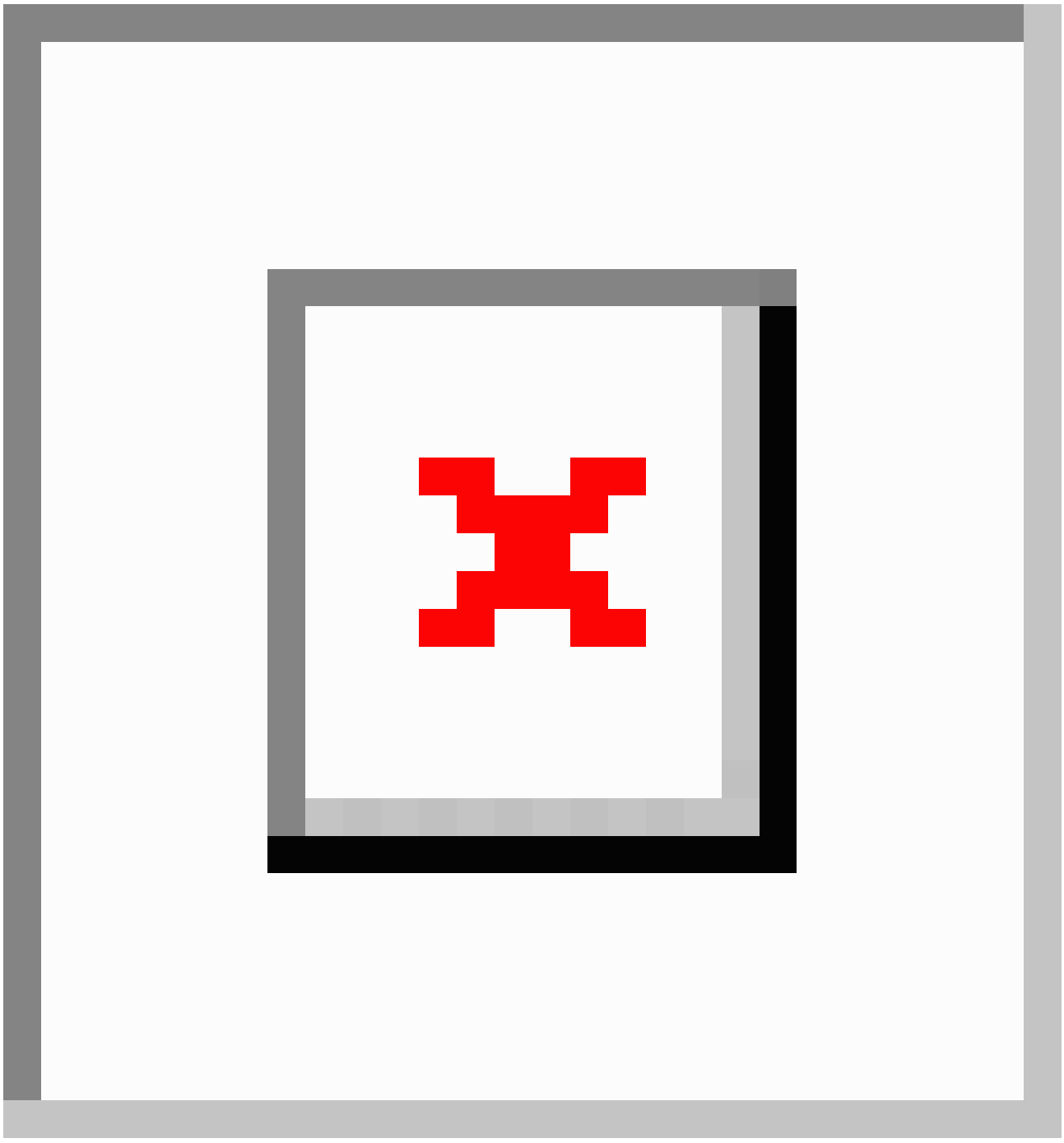
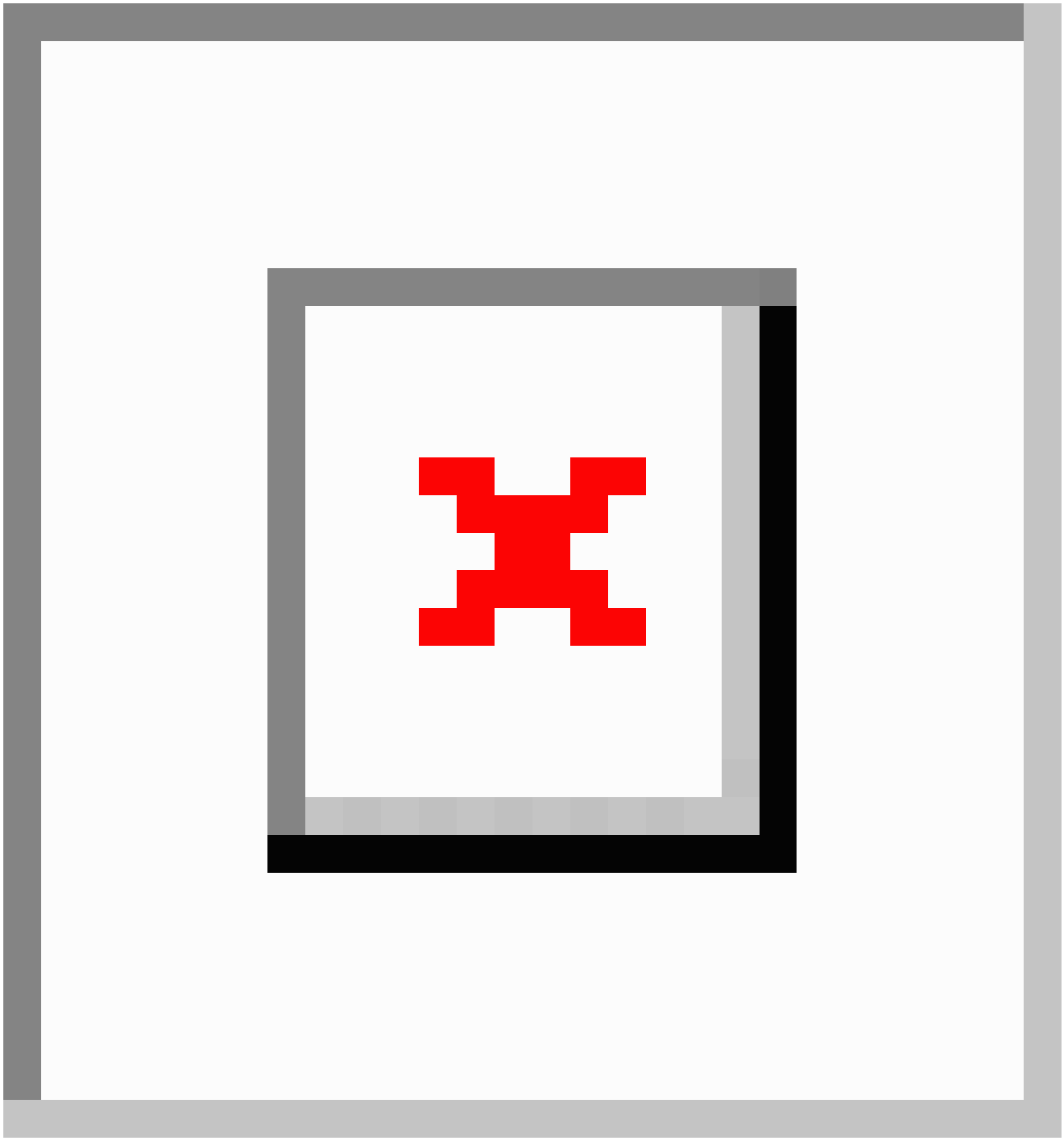
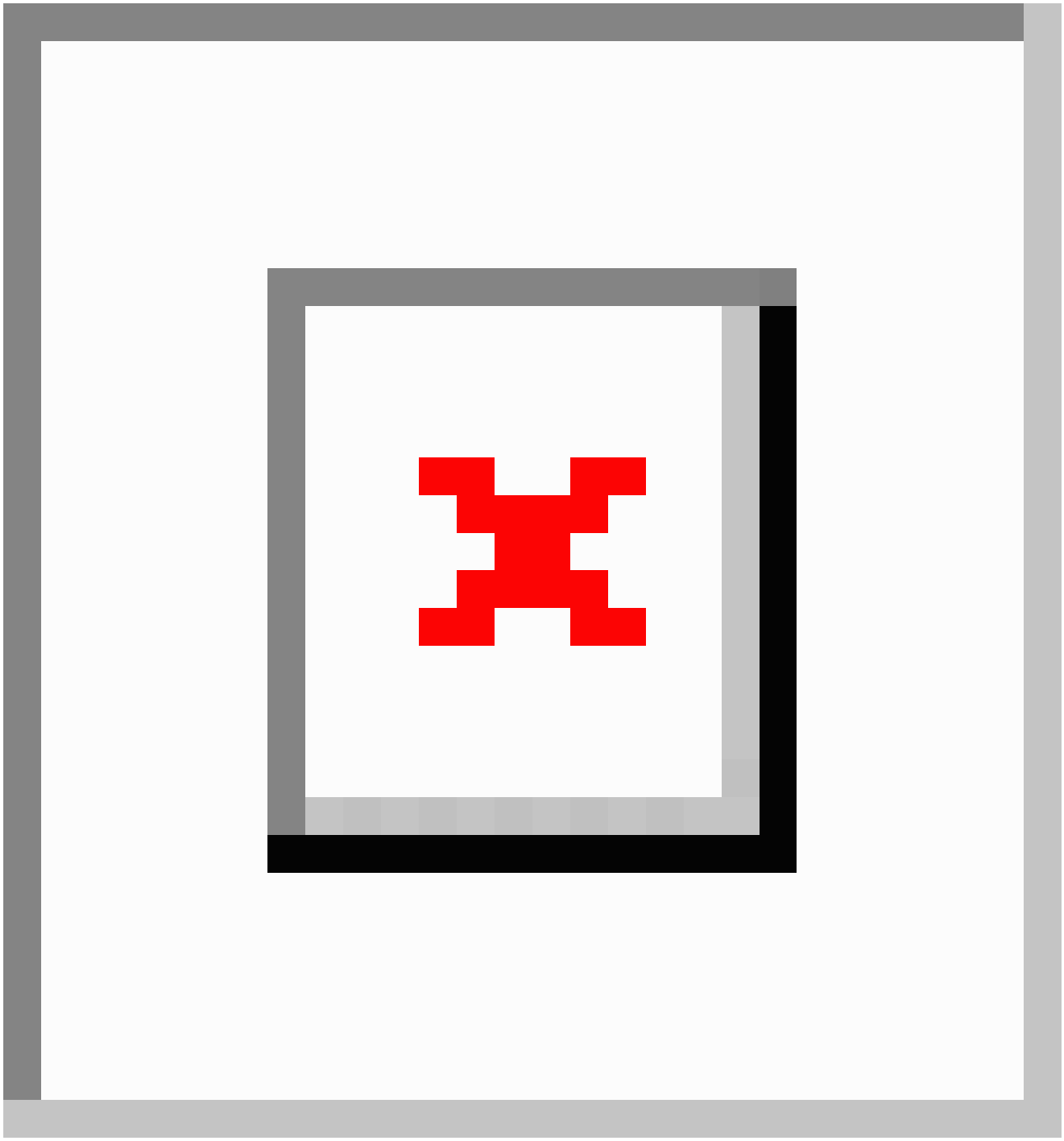


Figure 4. Individual baseline and week 6 APHAB scores. Thin lines represent each participant, and the thick line represents the group average. APHAB: Abbreviated Profile of Hearing Aids Benefit.



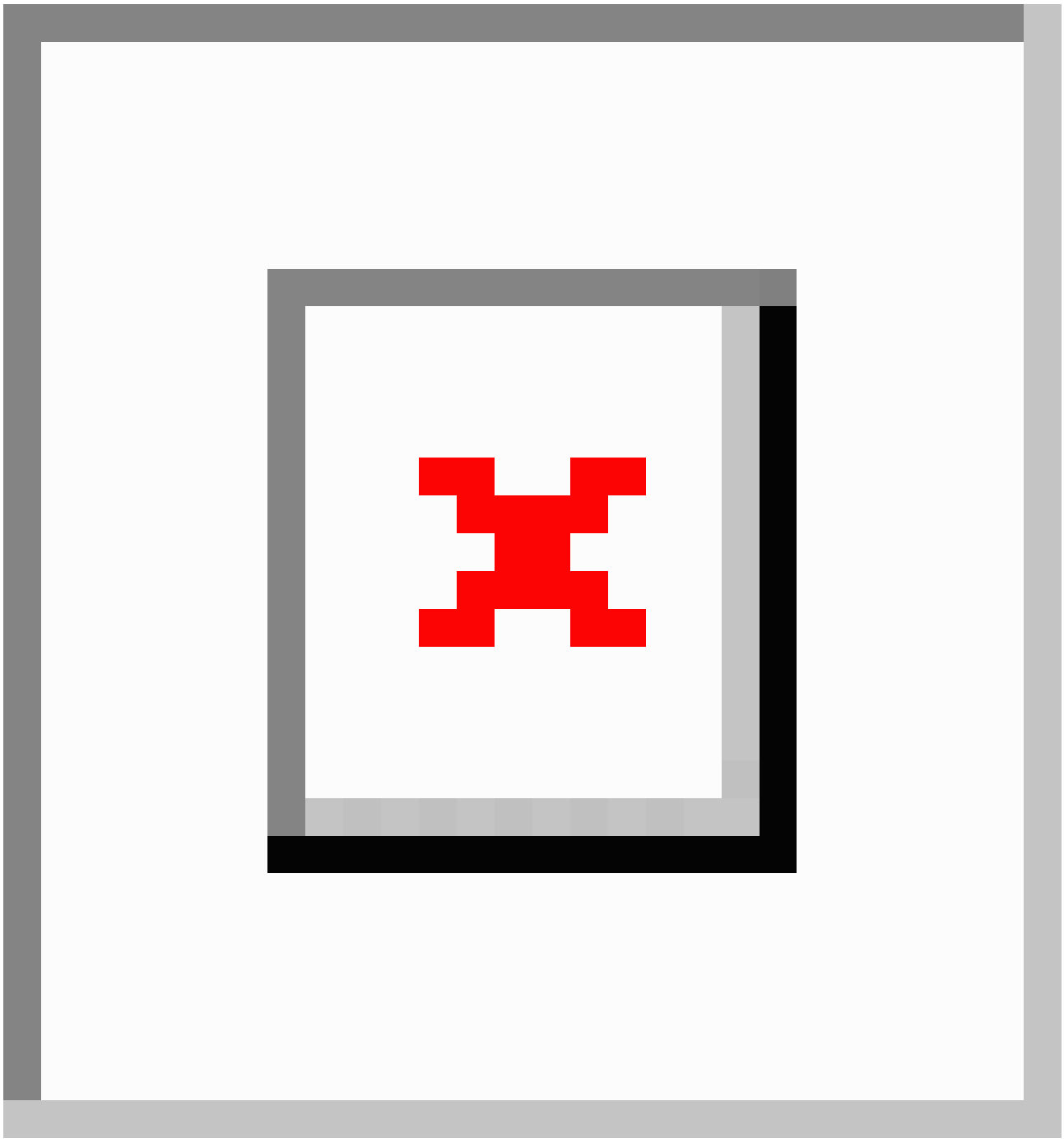
Time wearing the wristband and time exposed to speech were verified through the collection of data from backend logging that records when the wristband is turned on or off and when a phoneme is detected. As seen in Figure 5, participants wore the wristband for an average of 12.9 (SD 8.1) hours per day and were exposed to speech for an average of 6.7 (SD 3.3) hours per day.

Figure 5. Daily use for all participants. Bar height represents the total time the wristband was on. Orange represents the portion of time the wristband detected the presence of speech sounds. The dotted line represents the 1-hour minimum that participants were instructed to be around speech.



Simple linear regression analysis was used to test if a participant's baseline APHAB score explains their benefit APHAB score after 6 weeks, indicating that those with greater subjective difficulty understanding speech may stand to benefit the most from the haptic assistance of the wristband (Figure 6). The results of the regression indicate that the average baseline score explains 43% of the variation in the average APHAB benefit score at 6 weeks ($F_{1,14}$=10.55; P=.006). These results are significant at the P<.05 level.

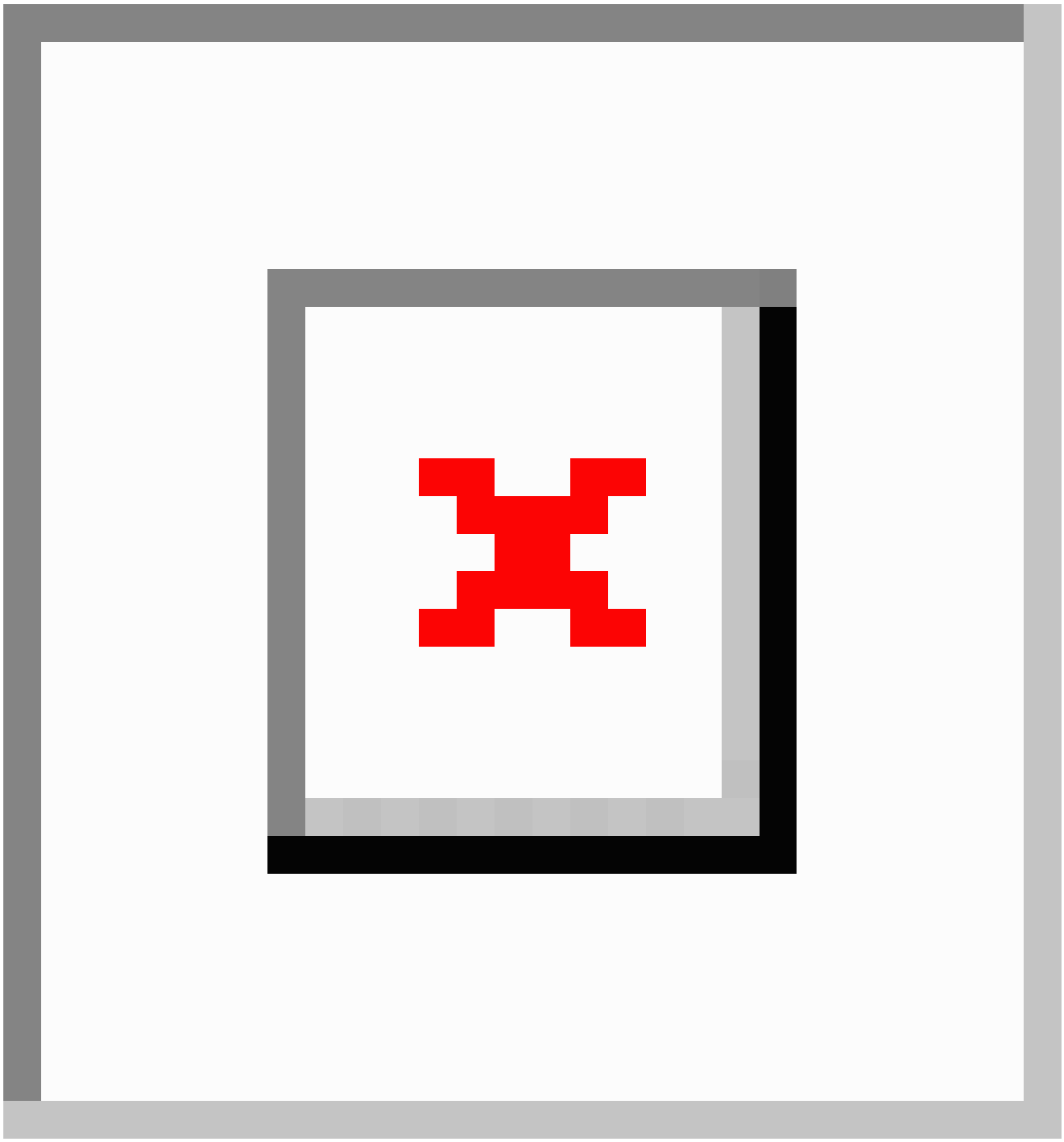
Figure 6. The baseline APHAB score correlates with the final APHAB benefit score. The linear regression demonstrates the correlation between the degree of disability without the assistance of Clarify at baseline and the final benefit score at week 6 with the aid of Clarify. APHAB: Abbreviated Profile of Hearing Aid Benefit.



We compared participants who used hearing aids to those who did not. A total of 9 participants used hearing aids to help them understand speech, and 7 of the participants did not. Results showed a 10.78 point greater APHAB benefit score at 6 weeks for participants who did not use hearing aids than for participants who did (t_{14}=2.14; P=.10, 2-tailed independent t test; Figure 7). While the difference in the benefit score between the two subgroups was not statistically significant, it did reach the 10-point threshold for clinical relevance [23,24]. The small sample size rendered the study underpowered to detect this difference at P<.05, and further study is necessary to validate this finding. Additionally, while the subgroup without hearing aids started the study at a higher level of disability, they ended the study at a lower level of disability than those with hearing aids. The subgroup without hearing aids started with a baseline APHAB score of 44.09 (SD 16.66) points, while the subgroup with hearing aids started with a baseline score of 37.40 (SD 14.61) points. The subgroup without hearing aids concluded the study with an APHAB score of 25.63 (SD 12.51) points, while the subgroup with hearing aids concluded the study with an APHAB score of 29.72 (SD 12.01) points. Another noteworthy difference between the subgroups was that the group who did not wear hearing aids demonstrated both a statistically significant and clinically meaningful aided APHAB benefit score from baseline, while the subgroup that did wear hearing

aids did not. The subgroup that did not wear hearing aids ended the study with an average APHAB benefit score from baseline of 18.45 points (SD 11.70 points; n=7; *P*=.005, 2-tailed dependent *t* test). The subgroup that wore hearing aids ended the study with an average APHAB benefit score from baseline of 7.67 points (SD 12.730 points; n=9; *P*=.11, 2-tailed dependent *t* test).

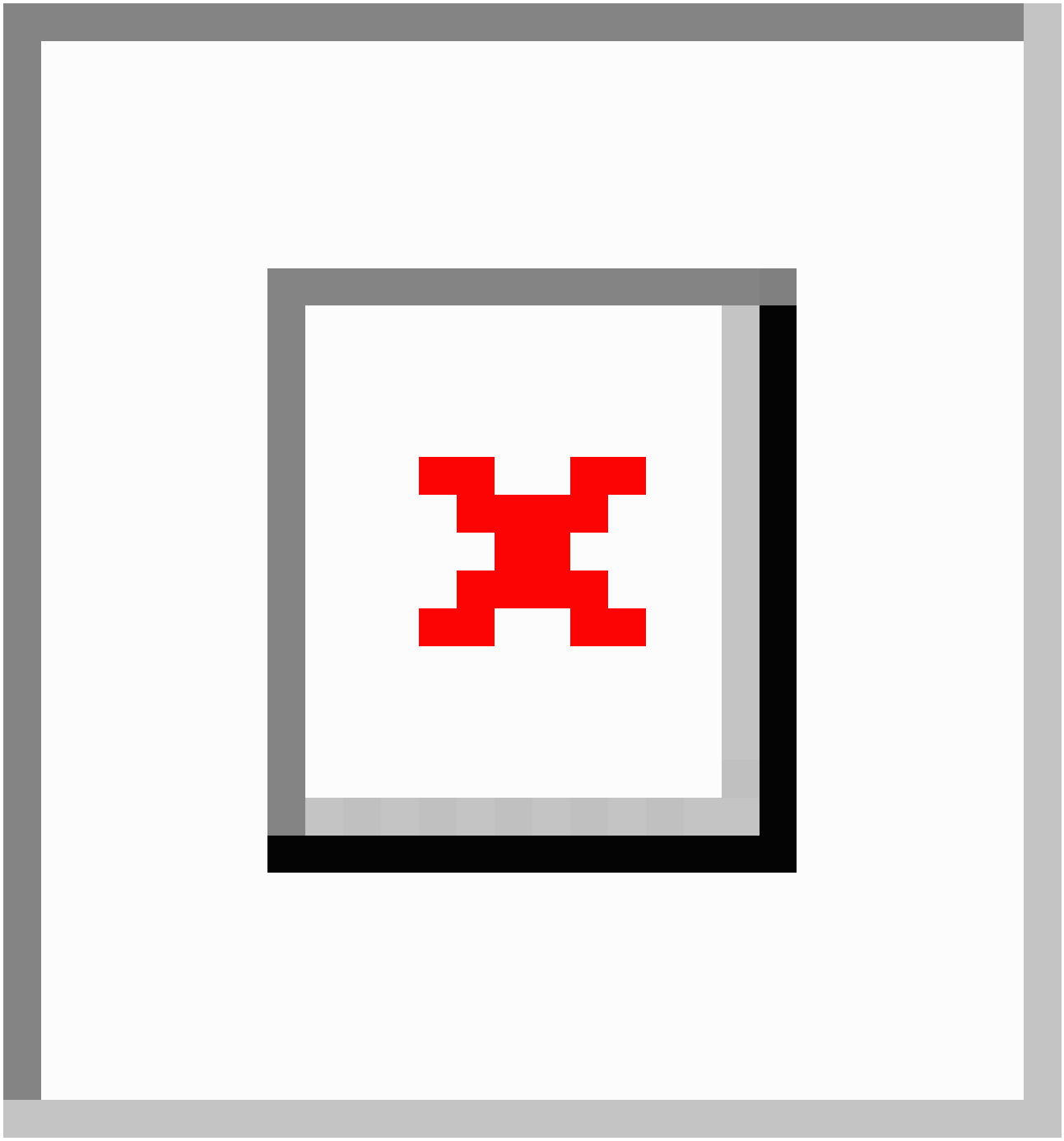
Figure 7. Non–hearing aid users ended the study with a higher benefit score than regular users of hearing aids. Error bars represent SE of the mean (SEM). Baseline SEM without hearing aids: 5.83. Baseline SEM with hearing aids: 4.59. Week 6 SEM without hearing aids: 4.38. Week six SEM with hearing aids: 3.78.



Subscale analyses were performed for ease of communication (EOC), BN, and reverberation (Figure 8 and Multimedia Appendix 3). These subscales are reflective of speech communication under ideal conditions, in noisy environments, and in reverberant environments [23]. The average benefit score for EOC was 15.44 (SD 13.88; N=16; *P*<.001, 2-tailed dependent *t* test). Those who wore hearing aids and those who did not wear hearing aids had similar EOC benefit scores (t_{14}=2.18; *P*=.60, 2-tailed independent *t* test). The average EOC benefit score for those with hearing aids was 13.57 (SD 15.71; n=9; *P*=.03, 2-tailed dependent *t* test), and the average EOC benefit score for those without hearing aids was 17.83 (SD 11.85; n=7; *P*=.01, 2-tailed dependent *t* test). The average benefit score for BN was 10.88 (SD 17.54; N=16; *P*=.03, 2-tailed dependent *t* test). The average BN benefit score for those without hearing aids was 16.99 points higher than those with hearing

aids (t_{14}=2.14; *P*=.05, 2-tailed independent *t* test). The average BN benefit score for those with hearing aids was 3.44 (SD 17.5; n=9; *P*=.54, 2-tailed dependent *t* test), and the average BN benefit score for those without hearing aids was 20.43 (SD 15.1; n=7; *P*=.01, 2-tailed dependent *t* test). The average benefit score for reverberation was 10.84 (SD 16.95; N=16; *P*=.02, 2-tailed dependent *t* test). The average reverberation benefit score for those without hearing aids was 11.12 points higher than those with hearing aids (t_{14}=2.14; *P*=.20, 2-tailed independent *t* test). The average reverberation benefit score for those without hearing aids was 17.10 (SD 16.0; n=7; *P*=.03, 2-tailed dependent *t* test), and the average reverberation benefit score for those with hearing aids was 5.98 (SD 17.0; n=9; *P*=.32, 2-tailed dependent *t* test).

Figure 8. APHAB subscale benefit scores at 6 weeks. Blue bars represent the entire participant group, orange bars are the subgroup who were regular users of hearing aids, and gray bars are the subgroup that did not wear hearing aids. There were 16 participants total, 9 who were regular users of hearing aids, and 7 who did not use hearing aids. Error bars represent the SE of the mean (SEM). Ease of communication SEM for total average: 3.47. Background noise SEM for total average: 4.38. Reverberation SEM for total average: 4.24. Ease of communication SEM for without hearing aids: 4.48. Background noise SEM for without hearing aids: 5.71. Reverberation SEM for without hearing aids: 6.03. Ease of communication SEM for with hearing aids: 5.24. Background noise SEM for with hearing aids: 5.41. Reverberation SEM for with hearing aids: 5.65. APHAB: Abbreviated Profile of Hearing Aid Benefit; HA: hearing aid.



Three of our participants requested to continue use of the wristband after the study ended, and hence, they did not fill out the final questionnaire. Of those who did, some had criticisms ("I'm really unsure if the Clarify band was helpful or not") and some had praise ("It was very beneficial. Thank you"); however, the comments were too few to be statistically meaningful.

Discussion

In this study, we expanded on our prior work that showed deaf and hard of hearing individuals are capable of identifying sound categories through patterns of vibration applied to the wrist [12]. Here, we demonstrated that individuals with high-frequency hearing loss can improve their subjective understanding of speech communication using vibrational representations of high-frequency speech sounds on the wrist. The results demonstrate that after 1 week of wearing the wristband, participants were able to improve their subjective ability to understand conversations during daily interactions. They then continued to improve, at a slower rate, throughout the 6-week study. This reflects prior research findings of an innate ability for those with hearing loss to rapidly learn to interpret tactile vibrations as a substitute for audio information [25]. The understanding of vibrations is further strengthened and perfected over time with practice as the portions of the auditory cortex that respond to tactile vibration expand [26-28].

We further found that participants who started the study with a higher baseline APHAB score experienced a greater improvement in their subjective ability to understand speech by the end of the 6-week trial. Of 16 participants, 14 ended the study with an APHAB score of 40 or below (which translates to perceived difficulty understanding speech less than half of the time). A total of 5 participants started the study with an unaided APHAB score of 50 points or higher; for 3 of them, the final APHAB benefit score was >30 points. One potential explanation for why participants who started the trial with greater difficulty understanding speech experienced greater improvement is that more of their auditory cortex was available for the interpretation of tactile sound representation [26]. It is also possible that participants who started the study with a higher APHAB score had more room for improvement, as higher APHAB scores indicate a higher degree of perceived disability. This could be an interesting topic for future research.

Participants without hearing aids demonstrated a trend toward higher self-reported benefit from vibrotactile sensory substitution for speech understanding, though this did not reach statistical significance. Given that this group started the study trending toward a higher APHAB score, we presume the difference is because the hearing aid group already benefits from their technology and therefore has less room for improvement. It is difficult to predict what the interaction between hearing aids and vibrotactile feedback will be because of the differing signal processing techniques used in digital hearing aid technologies. Digital hearing aids convert sound waves into numerical codes before amplifying them. This code contains information about a sound's frequency and amplitude, allowing the hearing aid to be specially programmed to amplify some frequencies more than others. Digital sound processing capabilities allow an audiologist to adjust the hearing aid to a user's needs and different listening environments. Digital hearing aids can also be programmed to focus on sounds coming from a specific direction. The wristband may represent sounds that differ significantly from those represented by the hearing aid. Future studies can explore directly connecting the wristband to the user's hearing aids through a Bluetooth signal so that the wristband's signals directly correspond with the sounds the user is hearing. For this study, the small sample size rendered the study underpowered to detect differences between those who used hearing aids and those who did not at $P<.05$. Future studies will be designed to investigate this finding further.

Individuals with hearing impairment have great difficulty understanding speech in the presence of BN. It is one of the primary complaints expressed by many with hearing loss, and one of the most difficult impairments to resolve. Individuals with hearing loss are unable to resolve the closely spaced harmonics of speech sounds to perform a spectral analysis with enough detail to extract the time-frequency portions of the speech that are relatively spared from corruption by the noise background [29]. In hearing aids, the BN modulators have not been shown to be highly effective at helping in these situations [30]. In this study, we demonstrated that the addition of vibrotactile feedback in the presence of BN enabled individuals who did not wear hearing aids to hear speech communication better based on their subjective experience (Figure 7). Interestingly, the final average BN score for the subgroup without hearing aids was 28.95 (SD 16.15; n=7) and the final average BN score for the subgroup with hearing aids was 40.04 (SD 18.78; n=9), suggesting that those who use hearing aids may benefit from using vibrotactile feedback during conversations with BN instead of using their hearing aids. While our data does not offer conclusive evidence of this due to several limitations, it does offer an area worth further exploration in larger studies.

Reverberation is the persistence of a sound after it is produced and is created when the sound is reflected off of surfaces or objects. It is most noticeable when the source of the sound has stopped, but the reflections continue. As the sound reflects off of surfaces and is absorbed by others, the quality of the sound degrades. Every room or outdoor environment has a different level of reverberation due to the construct of the room or area, the reflectiveness of the materials, and the objects in it. Reverberation is natural to every area, but in areas where the reverberation is very high, it can reduce speech intelligibility, especially when BN is also present. Individuals with hearing loss, including users of hearing aids, frequently report difficulty in understanding speech in reverberant, noisy situations [31]. Most hearing aids, both digital and analog, have limited ability to help individuals with hearing loss in areas of high reverberation [32]. We found that the addition of vibrotactile haptic vibration to the wrist in reverberant environments tended to help the participants without hearing aids more than those with hearing aids, though the difference did not reach statistical significance (Figure 6). One possibility to be tested is that individuals who use hearing aids may find haptic vibrations to be more helpful in reverberant environments when the hearing aids are removed because it would eliminate any conflict

between the digital processing of the hearing aid and the vibrational signals that are providing information about the sounds of speech without processing.

In the context of the APHAB, EOC describes the effort involved in communication under relatively easy listening environments. The interesting discovery from our results was that individuals who use hearing aids experienced a significant subjective improvement in their understanding of conversations under easy listening conditions. In easy listening environments where hearing aids help the most and perform the least amount of digital signal processing, the addition of haptic vibrations added the greatest amount of additional benefit. Upon completion of the trial, the average EOC score for the subset of participants who were users of hearing aids was 14.65 (SD 6.99; n=9), indicating little to no subjective difficulty understanding speech in easy listening environments. For the subset of participants who were not users of hearing aids, the average EOC score upon completion of the trial was 16.88 (7.73; n=7). Even without the additional help of hearing aids, these participants ended the study with an equivalent subjective capability for understanding speech in easier listening environments, despite starting the trial with a higher level of disability (Figure 8).

There are limitations to this study. First, the small sample size prevents extrapolation of the results to larger populations; this will be addressed in future studies. We were also limited in our ability to collect speech comprehension data in a noise-controlled environment with standardized volume controls—this is because the testing was done in participant homes instead of a laboratory. As a result, this study depended on self-report data (APHAB), which always has the potential to be influenced by a placebo effect. Another limitation is that some participant audiograms were assessed via phone apps rather than an audiologist's office; however, it should be noted that these appear to yield roughly equivalent results [5]. We also note that the specific type of hearing loss was not controlled beyond meeting the audiogram requirements. One final thing to note is that participants could move their hand (and, hence, their wristband), meaning that the microphone placement was not standardized in a single position. We do not consider this a limitation of the study, as the study is meant to test whether a vibrotactile wristband can be used to detect sound. The positive results reported here suggest that the mobility of the microphone does not present a problem.

We have demonstrated that vibrotactile sensory substitution helps individuals with high-frequency hearing loss improve their subjective understanding of verbal communication. The device demonstrated here is a wristband that delivers spatially distinguishable vibrations to the wrist in correspondence with high-frequency phonemes. We found that while both hearing aid and non–hearing aid users with high-frequency hearing loss reported a benefit, vibrotactile feedback tended to be more beneficial for non–hearing aid users. However, the small sample size rendered the study underpowered to detect this difference at $P<.05$, and further study is necessary to validate this finding. Finally, our results also demonstrated that those who started the study with a higher APHAB score (greater hearing disability) experienced the greatest amount of benefit from vibrotactile feedback.

Conflicts of Interest

DME is the chief executive officer of Neosensory, a neurotech company. IK and TF are employees of Neosensory, and MVP is a former employee of Neosensory.

Multimedia Appendix 1
Microphone characteristics.
[PDF File, 59 KB - xmed_v5i1e49969_app1.pdf]

Multimedia Appendix 2
Abbreviated Profile of Hearing Aids Benefit summary statistics per week for all participants, the subgroup that did not wear hearing aids, and the subgroup that did wear hearing aids. The benefit score is the baseline score minus the final score.
[PDF File, 83 KB - xmed_v5i1e49969_app2.pdf]

Multimedia Appendix 3
Abbreviated Profile of Hearing Aids Benefit subscale summary statistics per week for all participants, the subgroup that did not wear hearing aids, and the subgroup that did wear hearing aids.
[PDF File, 95 KB - xmed_v5i1e49969_app3.pdf]

References

1. Olusanya BO, Davis AC, Hoffman HJ. Hearing loss: rising prevalence and impact. Bull World Health Organ 2019 Oct 1;97(10):646-646A. [doi: 10.2471/BLT.19.224683] [Medline: 31656325]
2. Chang TY, Liu CS, Huang KH, Chen RY, Lai JS, Bao BY. High-frequency hearing loss, occupational noise exposure and hypertension: a cross-sectional study in male workers. Environ Health 2011 Apr 25;10:35. [doi: 10.1186/1476-069X-10-35] [Medline: 21518430]

3. Turner CW, Cummings KJ. Speech audibility for listeners with high-frequency hearing loss. Am J Audiol 1999 Jun;8(1):47-56. [doi: 10.1044/1059-0889(1999/002)] [Medline: 10499119]
4. Chen KH, Su SB, Chen KT. An overview of occupational noise-induced hearing loss among workers: epidemiology, pathogenesis, and preventive measures. Environ Health Prev Med 2020 Oct 31;25(1):65. [doi: 10.1186/s12199-020-00906-0] [Medline: 33129267]
5. Hong O, Kerr MJ, Poling GL, Dhar S. Understanding and preventing noise-induced hearing loss. Dis Mon 2013 Apr;59(4):110-118. [doi: 10.1016/j.disamonth.2013.01.002] [Medline: 23507351]
6. Michels TC, Duffy MT, Rogers DJ. Hearing loss in adults: differential diagnosis and treatment. Am Fam Physician 2019 Jul 15;100(2):98-108. [Medline: 31305044]
7. Jayakody DMP, Almeida OP, Speelman CP, et al. Association between speech and high-frequency hearing loss and depression, anxiety and stress in older adults. Maturitas 2018 Apr;110:86-91. [doi: 10.1016/j.maturitas.2018.02.002] [Medline: 29563040]
8. Feng Y, Yin S, Kiefte M, Wang J. Temporal resolution in regions of normal hearing and speech perception in noise for adults with sloping high-frequency hearing loss. Ear Hear 2010 Feb;31(1):115-125. [doi: 10.1097/AUD.0b013e3181bb69be] [Medline: 19816181]
9. Chung K. Challenges and recent developments in hearing AIDS. Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. Trends Amplif 2004;8(3):83-124. [doi: 10.1177/108471380400800302] [Medline: 15678225]
10. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. Int J Audiol 2014 Feb;53 Suppl 1:S18-S27. [doi: 10.3109/14992027.2013.860488] [Medline: 24447233]
11. Novich SD, Eagleman DM. Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. Exp Brain Res 2015 Oct;233(10):2777-2788. [doi: 10.1007/s00221-015-4346-1] [Medline: 26080756]
12. Perrotta MV, Asgeirsdottir T, Eagleman DM. Deciphering sounds through patterns of vibration on the skin. Neuroscience 2021 Mar 15;458:77-86. [doi: 10.1016/j.neuroscience.2021.01.008] [Medline: 33465416]
13. Eagleman DM, Perrotta MV. The future of sensory substitution, addition, and expansion via haptic devices. Front Hum Neurosci 2022 Jan 13;16:1055546. [doi: 10.3389/fnhum.2022.1055546] [Medline: 36712151]
14. Cieśla K, Wolak T, Lorens A, Mentzel M, Skarżyński H, Amedi A. Effects of training and using an audio-tactile sensory substitution device on speech-in-noise understanding. Sci Rep 2022 Feb 25;12(1):3206. [doi: 10.1038/s41598-022-06855-8] [Medline: 35217676]
15. Weisenberger JM, Russell AF. Comparison of two single-channel vibrotactile AIDS for the hearing-impaired. J Speech Hear Res 1989 Mar;32(1):83-92. [doi: 10.1044/jshr.3201.83] [Medline: 2704205]
16. Yesantharao LV, Donahue M, Smith A, Yan H, Agrawal Y. Virtual audiometric testing using smartphone mobile applications to detect hearing loss. Laryngoscope Investig Oto 2022 Sep 28;7(6):2002-2010. [doi: 10.1002/lio2.928] [Medline: 36544936]
17. Verrillo RT. Age related changes in the sensitivity to vibration. J Gerontol 1980 Mar;35(2):185-193. [doi: 10.1093/geronj/35.2.185] [Medline: 7410775]
18. Phatak SA, Yoon YS, Gooler DM, Allen JB. Consonant recognition loss in hearing impaired listeners. J Acoust Soc Am 2009 Nov;126(5):2683-2694. [doi: 10.1121/1.3238257] [Medline: 19894845]
19. Woods DL, Arbogast T, Doss Z, Younus M, Herron TJ, Yund EW. Aided and unaided speech perception by older hearing impaired listeners. PLoS One 2015 Mar 2;10(3):e0114922. [doi: 10.1371/journal.pone.0114922] [Medline: 25730423]
20. Sher AE, Owens E. Consonant confusions associated with hearing loss above 2000 Hz. J Speech Hear Res 1974 Dec;17(4):669-681. [doi: 10.1044/jshr.1704.669] [Medline: 4444287]
21. Mines MA, Hanson BF, Shoup JE. Frequency of occurrence of phonemes in conversational English. Lang Speech 1978 ;21(3):221-241. [doi: 10.1177/002383097802100302] [Medline: 732398]
22. Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: an ASR corpus based on public domain audio books. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): IEEE; 2015:5206-5210. [doi: 10.1109/ICASSP.2015.7178964]
23. Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. Ear Hear 1995 Apr;16(2):176-186. [doi: 10.1097/00003446-199504000-00005] [Medline: 7789669]
24. Cox RM. Administration and application of the APHAB. Hearing J 1997 Apr;50(4):32. [doi: 10.1097/00025572-199704000-00002]
25. Soto-Faraco S, Deco G. Multisensory contributions to the perception of vibrotactile events. Behav Brain Res 2009 Jan 23;196(2):145-154. [doi: 10.1016/j.bbr.2008.09.018] [Medline: 18930769]
26. Auer ET, Bernstein LE, Sungkarat W, Singh M. Vibrotactile activation of the auditory cortices in deaf versus hearing adults. Neuroreport 2007 May 7;18(7):645-648. [doi: 10.1097/WNR.0b013e3280d943b9] [Medline: 17426591]
27. Good A, Reed MJ, Russo FA. Compensatory plasticity in the deaf brain: effects on perception of music. Brain Sci 2014 Oct 28;4(4):560-574. [doi: 10.3390/brainsci4040560] [Medline: 25354235]
28. Levänen S, Jousmäki V, Hari R. Vibration-induced auditory-cortex activation in a congenitally deaf adult. Curr Biol 1998 Jul 16;8(15):869-872. [doi: 10.1016/S0960-9822(07)00348-X] [Medline: 9705933]

29. McArdle R, Wilson RH. Speech perception in noise: the basics. Perspect Hear Hear Disord Res Diagnostics 2009 Feb 1;13(1):4-13 [FREE Full text]
30. Healy EW, Yoho SE. Difficulty understanding speech in noise by the hearing impaired: underlying causes and technological solutions. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:89-92. [doi: 10.1109/EMBC.2016.7590647] [Medline: 28268288]
31. Cueille R, Lavandier M, Grimault N. Effects of reverberation on speech intelligibility in noise for hearing-impaired listeners. R Soc Open Sci 2022 Aug 31;9(8):210342. [doi: 10.1098/rsos.210342] [Medline: 36061524]
32. Reinhart PN, Souza PE, Srinivasan NK, Gallun FJ. Effects of reverberation and compression on consonant identification in individuals with hearing impairment. Ear Hear Mar-Apr 2016;37(2):144-152. [doi: 10.1097/AUD.0000000000000229] [Medline: 26462171]

Abbreviations

APHAB: Abbreviated Profile of Hearing Aid Benefit
BN: background noise
EOC: ease of communication
GRMS: root mean squared acceleration from gravity
LRA: linear resonant actuator

Edited by E Meinert; submitted 14.06.23; peer-reviewed by Anonymous, Anonymous, R Eikelboom; revised version received 23.11.23; accepted 13.12.23; published 09.02.24.

Please cite as:
Kohler I, Perrotta MV, Ferreira T, Eagleman DM
Cross-Modal Sensory Boosting to Improve High-Frequency Hearing Loss: Device Development and Validation
JMIRx Med 2024;5:e49969
URL: https://xmed.jmir.org/2024/1/e49969
doi:10.2196/49969

COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study

Troy B Puga[1,*], BSc; Joshua Schafer[2,*], BSc; Grace Thiel[1,*], BSc; Nicholas Scigliano[1,*], BSc; Tiffany Ruan[1,*], BSc; Andres Toledo[1,*], BSc; Prince N Agbedanu[3,*], BSc, PhD; Kevin Treffer[4,*], BSc, DO

1

2

3

4

[*]all authors contributed equally

Corresponding Author:
Troy B Puga, BSc

Related Articles:

Companion article: https://preprints.jmir.org/preprint/45688

Companion article: https://med.jmirx.org/2024/1/e55900

Companion article: https://med.jmirx.org/2024/1/e55997

Companion article: https://med.jmirx.org/2024/1/e56040

Companion article: https://med.jmirx.org/2024/1/e56169

Companion article: https://med.jmirx.org/2024/1/e55863

Abstract

Background: In 2020, COVID-19 spread across the world and brought normal daily life to a halt, causing the shutdown of nearly everything in order to prevent its spread. The National Football League (NFL) similarly experienced shutdowns and the resulting effects, leaving athletes unable to train in some of the most advanced facilities with many of the best trainers in the world. A previous study, titled *COVID-19 Return to Sport: NFL Injury Prevalence Analysis*, determined that there was increased injury prevalence during the 2020 season, likely due to decreased physiological adaptations within athletes' bodies as a result of facility shutdowns. Understanding injury epidemiology is vital to the prevention of injuries and the development of return-to-play protocols.

Objective: The objective of this study is to perform a follow-up study to *COVID-19 Return to Sport: NFL Injury Prevalence Analysis* in order to examine the longitudinal effects of the COVID-19 pandemic on injury epidemiology. This study examines if there was a recovery to baseline levels of injuries or if there are still lingering effects from the COVID-19 pandemic–induced spike in injuries.

Methods: To determine if there was change in the number of injuries for each season, injury tallies collected from the 17-week-long 2018, 2019, and 2020 NFL regular seasons were compared with those from the 18-week-long 2021 and 2022 NFL regular seasons. A Kruskall-Wallis test with post hoc Dunn analysis was conducted to compare the rate of injuries per team per week between each of the 2018, 2019, 2020, 2021, and 2022 regular seasons.

Results: The Kruskall-Wallis test revealed an *H* statistic of 32.61 ($P<.001$) for the comparison of the injury rates across the 5 seasons. The post hoc Dunn analysis showed that 2020 had a statistically significant difference when compared with each of the 2018 ($P<.001$), 2019 ($P=.04$), 2021 ($P=.02$), and 2022 ($P=.048$) seasons. The 2019 season showed no statistical significance when compared with the 2021 ($P=.23$) and 2022 ($P=.13$) seasons.

Conclusions: The results of this follow-up study, combined with the previous study, show that extended training interruptions stemming from COVID-19 in 2020 induced detraining and led to increased injuries. Additionally, the results of this study show that retraining can occur, resulting in the development of injury protective factors, as injury rates returned to baseline levels after

2020. This is the first large-scale and long-term opportunity to demonstrate the effects of these principles and how they are important to understanding injury epidemiology.

(JMIRx Med 2024;5:e45688) doi:10.2196/45688

KEYWORDS

COVID-19; injury; prevalence; adaptation; sports medicine; follow-up; training; football; epidemiology; sport; athlete; athletic; injuries

Introduction

In 2020, the COVID-19 pandemic spread across the world and led to the shutdown of everything except essential services. Sports were no exception, as competition came to an immediate halt. The National Football League (NFL) was one such sport that was shutdown, as they closed all facilities from late March 2020 until late May 2020 [1,2]. While facilities opened back up in May, players were unable to return until the end of July [3]. This shutdown eliminated most of the training period, giving players a narrow 20-day period to reacclimate before the start of the 2020 season [3]. A previous study, titled *COVID-19 Return to Sport: NFL Injury Prevalence Analysis*, examined the effects of this shutdown on injury epidemiology [2]. This study showed that there was increased injury prevalence during the 2020 NFL season, which was impacted by the COVID-19 pandemic [2]. The study concluded that the increased injury prevalence was due to a decline in athlete training as a result of the shutdowns [2]. The decline in training, and hence, the lack of bodily preparation for the 2020 NFL season, resulted in decreased physiological adaptations of players' bodies to strenuous exertions, which is a hallmark of the sport [2]. This study serves as a follow-up study to examine the longitudinal effects of the COVID-19 pandemic shutdowns on injury prevalence in the NFL following the 2020 NFL season.

The NFL is an American football league that is composed of 32 teams with some of the best athletes in the world. The NFL is widely recognized as having some of the best trainers and facilities to ensure that these high-level athletes are prepared to play each season. NFL athletes follow intense training in order to physically prepare their bodies for the demand of a strenuous season [4]. Although training may not prevent every injury, research has shown that training produces a protective effect from injury [5-7]. Traditionally, the NFL was played over a 17-week-long NFL season each year until the 2021 season, when they transitioned to an 18-week-long NFL season for all future seasons [8].

Previous research demonstrated that the COVID-19 pandemic had acute effects on injury prevalence during the 2020 NFL season [2]. Much of this can be attributed to the decreased physiological adaptations that occurred during shutdowns, when athletes were unable to access NFL training facilities [1,2]. It is important to note that although teams could provide players with up to US $1500 worth of at-home training equipment, the players still lacked access to on-site athletic trainers and recovery facilities and were unable to partake in normal preseason training [9]. Any at-home workouts were considered voluntary [9]. Therefore, these athletes most likely exhibited the effects of detraining, a process in which athletes undergo a loss of physiological and performance-based adaptations from previous training due to a lack of sufficient training [3,10]. It is believed that many athletes across the United States and the world exhibited the effects of acute detraining during the COVID-19 shutdown [3,11]. Previous shutdowns in the NFL have shown that detraining can occur, such as during the 2011 lockout, where the preseason saw an increased number of Achilles tendon injuries [12]. Training is one of the most important interventions for injury prevention and improved athletic performance in athletes of all populations [5-7,13-15].

The effects of detraining are a serious concern in regard to injury epidemiology. We can clearly see that detraining has significant potential to lead to injuries; however, complete recovery from detraining can be obtained through a sufficient training program [3,16]. The amount of time necessary to recover from the effects of detraining is widely debated [3]. The amount of time needed for training adaptations often differs between people based on training levels, genetics, and a host of other factors [3,17-19]. Many of these genetic and environmental factors contribute to the variability of injury recovery times and why some athletes may recover more quickly [20-22]. Although we know that recovery is complex and difficult to predict, research has suggested that the time necessary to fully return from detraining is often similar to or longer than the length of detraining [16]. The variations in training adaptations and recovery from detraining makes the examination of the longitudinal effects of the COVID-19 pandemic that much more important to understanding longitudinal injury epidemiology.

Longitudinal research on injury epidemiology due to a major event, such as the COVID-19 pandemic, is scarce. The goal of this research is to understand the effects of the COVID-19 pandemic on longitudinal injury epidemiology. The authors hypothesized that injury prevalence for the 2021 and 2022 NFL seasons would be lower than that of the 2020 NFL season, which was affected by the COVID-19 pandemic. Correspondingly, it was hypothesized that decreased injury prevalence would result from athletes having full access to sports performance facilities and staff to properly train their bodies during the 2021 and 2022 NFL seasons. It is believed that during these 2 seasons, athletes had adequate time and preparation to induce the necessary physiological changes and undergo retraining, thus avoiding the effects of detraining observed following the COVID-19 pandemic–induced shutdowns.

Methods

Ethical Considerations

This study did not require institutional review board approval as all data are publicly available and this study does not qualify

as human subjects research per US Department of Health and Human Services policy 45 CFR 46.102.f.

Study Design

The methodology for this study design was adapted from the previous study, *COVID-19 Return to Sport: NFL Injury Prevalence Analysis* [2]. Data for the 2020 NFL regular season were collected in the previous study and used for further analysis [2]. The number of injuries was tallied for the 18-week-long 2021 and 2022 NFL regular seasons using the weekly published injury reports by each NFL team. Injury reports are made publicly available by each team in the NFL. If an official injury report was not available through the individual team media, a deferment was made to the official NFL website. Per the NFL injury report policy, teams are required to report injuries throughout the week of a game, so any and all information should be considered accurate [23]. Athletes listed with the same injury for consecutive weeks were only counted once in order to prevent any repeat data. If an athlete presented with an injury to a different anatomical region, this was counted as a new unique injury. No data were repeated at any point during the collection process. Contact injuries were included in this study as this is a nonmodifiable risk factor that cannot be controlled due to football being a contact sport [2,24]. COVID-19 infection, sick days, and nonmedical days off were not included in the injury tally. Illnesses were not included in this study, because illnesses are not considered a physical injury and should be reported separately from injuries when performing injury epidemiological studies [2,25]. All other soft tissue injuries and concussions were included during the first week of the associated injury report. The 2020 season, which was played after the COVID-19–induced shutdowns, was used for comparison with the subsequent 2021 and 2022 seasons.

Data Analysis

The data analysis for this study was conducted differently from the previous study [2], as there is a need to correct for the change from the 17-week-long NFL seasons in 2018, 2019, and 2022 to the 18-week-long NFL seasons in 2021 and 2022 [8]. The comparison required first dividing the number of injuries per team by the number of weeks for each season. This adjustment allowed for the comparison of 2 different regular season lengths. A Kruskall-Wallis test with post hoc Dunn analysis was conducted on the number of injuries per team per week rate to compare the 2018, 2019, 2020, 2021, and 2022 seasons.

Results

The 2018 season had a total of 1561 injuries (Figures 1 and 2) [2]. This number was divided by 17, the number of weeks in the NFL regular season for the 2018, 2019, and 2020 seasons, to produce a rate that allowed for an appropriate comparison with the change to 18-week seasons in 2021 and 2022. Thus, the 2018 season produced a rate of 91.8 injuries per week and a rate of 2.9 injuries per team per week. The 2019 season had a total of 1897 injuries [2], producing a league-wide rate of 111.6 injuries per week and a rate of 3.5 injuries per team per week. The 2020 season had a total of 2484 injuries, as indicated in the previous study [2]. It produced a league-wide rate of 146.1 injuries per week and a rate of 4.6 injuries per team per week. The 2021 season had a total of 2210 injuries, which was divided by 18 to correspond with the number of weeks in the 2021 regular season. This produced a league-wide rate of 122.8 injuries per week and a rate of 3.8 injuries per team per week. The 2022 season had a total of 2257 injuries. This was also divided by 18 to represent the number of weeks in the 2022 regular season. This produced a league-wide rate of 125.4 injuries per week and a rate of 3.3 injuries per team per week.

Data analysis with a Kruskall-Wallis test found that the rate of injuries per team per week differed between the 5 seasons with an H statistic of 32.61 ($P<.001$). From the post hoc Dunn analysis, the rate of injuries per team per week for the 2018 NFL season was statistically significantly different than those of the 2019 ($P=.04$), 2020 ($P<.001$), 2021 ($P<.001$), and 2022 ($P<.001$) seasons. The rate of injuries per team per week for the 2019 NFL season was statistically significantly different than that of the 2020 ($P=.04$) season, but there was no significant difference when compared with those of the 2021 ($P=.23$) and 2022 ($P=.13$) seasons. The rate of injuries per team per week for the 2020 COVID-19–impacted NFL season was statistically significantly different than those of the 2021 ($P=.02$) and 2022 ($P=.048$) seasons. Comparison of the rate of injuries per team per week of the 2021 and 2022 seasons did not produce a statistically significant difference ($P=.76$).

Figure 1. The total number of injuries per NFL regular season, including the 17-week-long 2018, 2019, and 2020 NFL seasons from the previous study [2] and the 18-week-long 2021 and 2022 NFL seasons. Error bars signify SE. NFL: National Football League.

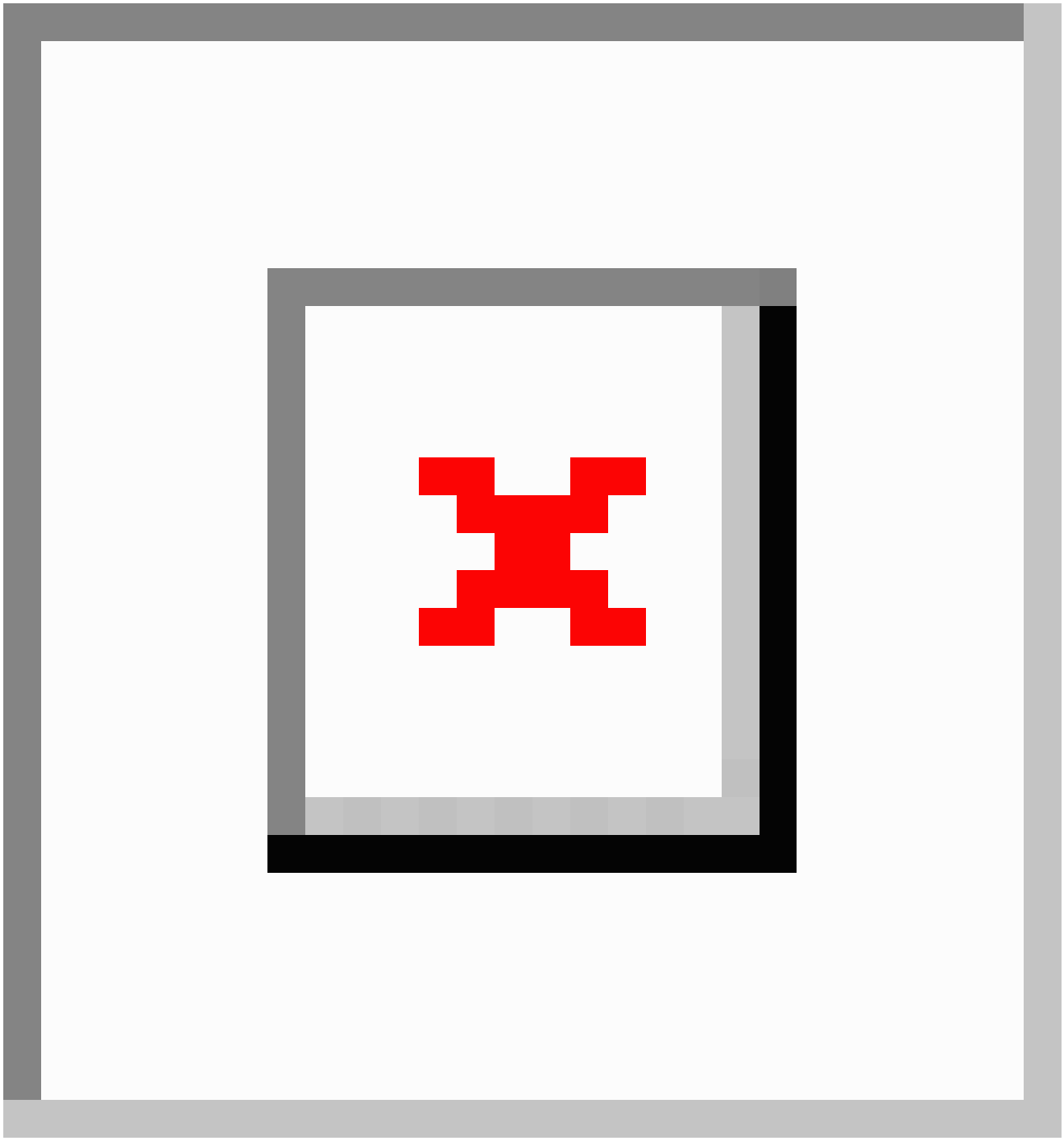
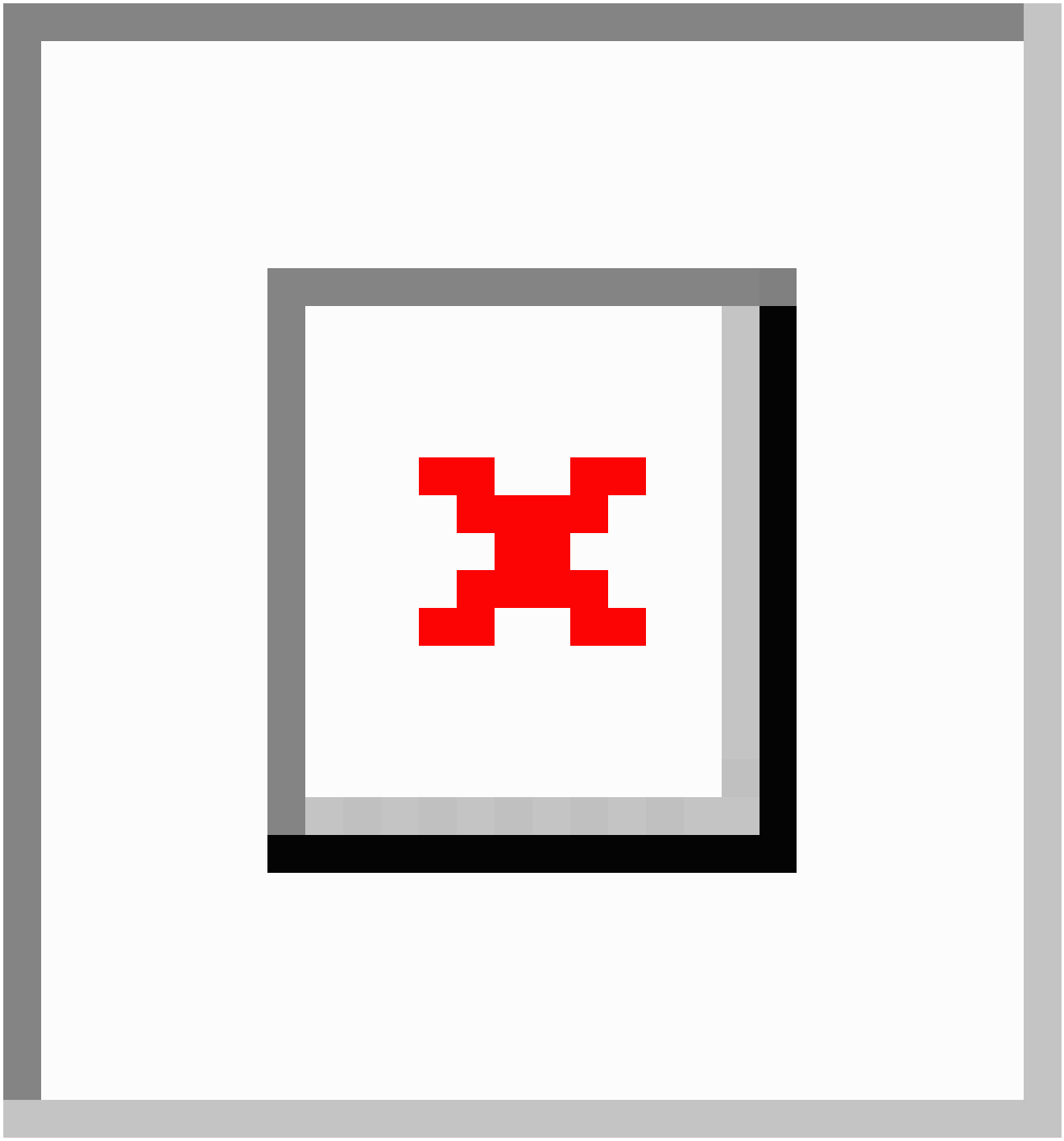


Figure 2. Number of injuries per week per team for each NFL regular season from 2018 - 2022. Error bars signify SE. NFL: National Football League.



Discussion

The study results showed statistically significant differences when comparing the 2020 COVID-19–impacted NFL season with the previous 2018 and 2019 seasons, as well as when compared with the subsequent 2021 and 2022 NFL seasons. The study also showed that there was a lower total number of injuries across the league in 2021 (2210 injuries) and 2022 (2257 injuries) when compared with the 2020 (2484 injuries) COVID-19–impacted NFL season. This occurred even with the addition of an extra week of regular season play in the 2021 and 2022 seasons when compared with the 2020 season [8]. The 2021 and 2022 seasons showed no statistically significant difference when compared with the 2019 season, indicating that injury levels returned to a similar level after 2020. This shows that there was a clear epidemiological spike in 2020. The previous study showed the impact of the COVID-19 pandemic, which led to a higher number of injuries in the 2020 NFL season when compared with earlier seasons—likely due to decreased training adaptations. This was thought to have stemmed from facility lockdowns aimed at mitigating the spread of COVID-19 [2]. This study demonstrated recovery from that spike back to prepandemic season levels, indicating that retraining and recovery from detraining is possible.

Previous research has demonstrated the effects of shutdowns and detraining, which lead to a negative impact on athletic performance and health [2,3,10-12]. Preparation for athletic competition is vital for the health and performance of athletes

[5-7,13-15]. This follow-up study aimed to examine if the COVID-19 pandemic's effects on injury epidemiology persisted or if athletes were able to return their bodies back to peak prepandemic shape. The decline in injuries after the COVID-19 pandemic season, as demonstrated in this follow-up study, showed that athletes were able to return to prepandemic athletic conditioning. This was most likely achieved through having a full offseason with full-access training for the 2021 and 2022 seasons. NFL players train in some of the best facilities in the world while under the supervision of renowned sports medicine personnel. It can be concluded from this follow-up study that the COVID-19 pandemic did have an acute effect on injuries, with a spike in the 2020 season. However, the COVID-19 pandemic did not have a longitudinal effect on injuries, as demonstrated by the decrease in injuries during the 2021 and 2022 regular seasons when compared with 2020. This follow-up study further suggests that athletic preparation through performance training is vital for the preparation of sport and injury prevention.

However, this study is not without limitations, and these limitations are consistent with the previous study. Similar to the previous study, this study is limited by the potential underreporting of injuries by players, unbeknownst to team personnel [2]. Another potential limitation, consistent with the previous study, is that the exact difference in training hours between seasons is unknown and variable by team to some degree [2]. However, it is known that there were limitations for training in 2020 [1] and that any survey administered about training limitations would be limited due to recall bias [2]. It is also possible that other factors influenced the findings of this study. However, the span of years included would hopefully account for any other possible differences besides the COVID-19–induced lockdown. The authors believe that despite these limitations, this study provides an accurate representation of the effects of detraining on injury epidemiology.

This research has demonstrated the importance of performance training for sport on injury epidemiology. This study shows that recovery from detraining is possible under the proper training conditions, in which athletes induce adaptation and preparation for sport. This is the first large-scale opportunity available to study detraining and retraining principles from a long-term perspective. Findings from this study serve as a foundational piece of research in proving what many have been hypothesized in regard to training and injury prevention.

Further studies must determine exactly how much time is needed to return to sport from a major event such as the COVID-19 pandemic, injuries, or time off. Qualified performance coaches and rehabilitation professionals are vital to help solve these issues. Further research must also be done at the collegiate and amateur levels to examine the effects of detraining and retraining, as these individuals may not have access to the care and facilities that NFL players receive. The findings from the previous study and this follow-up study serve as a starting point to future discussions of the necessary preparation times and levels. Further injury epidemiological trials and observational studies must be conducted to continue the evolution of return-to-sport protocols and training programs to combat detraining in adverse situations.

Acknowledgments

We would like to thank Elijah Theye, MPH, and the University of Nebraska Medical Center College of Public Health regarding their help with the biostatistical analysis of our results.

Conflicts of Interest

None declared.

References

1. Barrabi T. NFL facilities can reopen from coronavirus shutdown starting May 19, Goodell says. Fox Business. 2020 May 15. URL: https://www.foxbusiness.com/sports/nfl-facilities-reopen-date-coronavirus-roger-goodell [accessed 2022-06-06]
2. Puga TB, Schafer J, Agbedanu PN, Treffer K. COVID-19 return to sport: NFL injury prevalence analysis. JMIRx Med 2022 Apr;3(2):e35862. [doi: 10.2196/35862] [Medline: 35511457]
3. Girardi M, Casolo A, Nuccio S, Gattoni C, Capelli C. Detraining effects prevention: a new rising challenge for athletes. Front Physiol 2020 Oct 15;11:588784. [doi: 10.3389/fphys.2020.588784] [Medline: 33178050]
4. Ebben WP, Blackard DO. Strength and conditioning practices of National Football League strength and conditioning coaches. J Strength Cond Res 2001 Feb;15(1):48-58. [Medline: 11708706]
5. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? Br J Sports Med 2016 Mar;50(5):273-280. [doi: 10.1136/bjsports-2015-095788] [Medline: 26758673]
6. Lauersen JB, Bertelsen DM, Andersen LB. The effectiveness of exercise interventions to prevent sports injuries: a systematic review and meta-analysis of randomised controlled trials. Br J Sports Med 2014 Jun;48(11):871-877. [doi: 10.1136/bjsports-2013-092538] [Medline: 24100287]
7. Lauersen JB, Andersen TE, Andersen LB. Strength training as superior, dose-dependent and safe prevention of acute and overuse sports injuries: a systematic review, qualitative analysis and meta-analysis. Br J Sports Med 2018 Dec;52(24):1557-1563. [doi: 10.1136/bjsports-2018-099078] [Medline: 30131332]

8. Graziano D. NFL moves to 17-game regular season in 2021: what it means for teams, players, revenue, and fans. ESPN. 2021 Mar 30. URL: https://www.espn.com/nfl/story/_/id/31159686/nfl-moves-17-game-regular-season-2021-means-teams-players-revenue-fans [accessed 2022-06-14]
9. Battista J. Virtual offseason program set to begin April 20. NFL. 2020 Apr 13. URL: https://www.nfl.com/news/virtual-offseason-program-set-to-begin-april-20-0ap3000001109298 [accessed 2023-07-18]
10. Mujika I, Padilla S. Detraining: loss of training-induced physiological and performance adaptations. part I: short term insufficient training stimulus. Sports Med 2000 Aug;30(2):79-87. [doi: 10.2165/00007256-200030020-00002] [Medline: 10966148]
11. Nakisa N, Rahbardar MG. Evaluating the probable effects of the COVID-19 epidemic detraining on athletes' physiological traits and performance. Apunts Sports Medicine 2021 ;56(211):100359. [doi: 10.1016/j.apunsm.2021.100359]
12. Myer GD, Faigenbaum AD, Cherny CE, Heidt RSJ, Hewett TE. Did the NFL lockout expose the Achilles heel of competitive sports? J Orthop Sports Phys Ther 2011 Oct;41(10):702-705. [doi: 10.2519/jospt.2011.0107] [Medline: 21941038]
13. Harries SK, Lubans DR, Callister R. Resistance training to improve power and sports performance in adolescent athletes: a systematic review and meta-analysis. J Sci Med Sport 2012 Nov;15(6):532-540. [doi: 10.1016/j.jsams.2012.02.005] [Medline: 22541990]
14. Kraemer WJ, Ratamess NA, French DN. Resistance training for health and performance. Curr Sports Med Rep 2002 Jun;1(3):165-171. [doi: 10.1249/00149619-200206000-00007] [Medline: 12831709]
15. Øvretveit K, Tøien T. Maximal strength training improves strength performance in grapplers. J Strength Cond Res 2018 Dec;32(12):3326-3332. [doi: 10.1519/JSC.0000000000002863] [Medline: 30247275]
16. Joo CH. The effects of short term detraining and retraining on physical fitness in elite soccer players. PLoS One 2018 May 10;13(5):e0196212. [doi: 10.1371/journal.pone.0196212] [Medline: 29746505]
17. Vikmoen O, Raastad T, Ellefsen S, Rønnestad BR. Adaptations to strength training differ between endurance-trained and untrained women. Eur J Appl Physiol 2020 Jul;120(7):1541-1549. [doi: 10.1007/s00421-020-04381-x] [Medline: 32372216]
18. Bouchard C, Rankinen T, Timmons JA. Genomics and genetics in the biology of adaptation to exercise. Compr Physiol 2011 Jul;1(3):1603-1648. [doi: 10.1002/cphy.c100059] [Medline: 23733655]
19. Mann TN, Lamberts RP, Lambert MI. High responders and low responders: factors associated with individual variation in response to standardized training. Sports Med 2014 Aug;44(8):1113-1124. [doi: 10.1007/s40279-014-0197-3] [Medline: 24807838]
20. Lim T, Santiago C, Pareja-Galeano H, et al. Genetic variations associated with non-contact muscle injuries in sport: a systematic review. Scand J Med Sci Sports 2021 Nov;31(11):2014-2032. [doi: 10.1111/sms.14020] [Medline: 34270833]
21. Booth-Kewley S, Schmied EA, Highfill-McRoy RM, Sander TC, Blivin SJ, Garland CF. A prospective study of factors affecting recovery from musculoskeletal injuries. J Occup Rehabil 2014 Jun;24(2):287-296. [doi: 10.1007/s10926-013-9456-7] [Medline: 23797182]
22. Kvist J. Rehabilitation following anterior cruciate ligament injury: current recommendations for sports participation. Sports Med 2004;34(4):269-280. [doi: 10.2165/00007256-200434040-00006] [Medline: 15049718]
23. 2017 personel (injury) report policy. NFL. 2017. URL: https://operations.nfl.com/media/2683/2017-nfl-injury-report-policy.pdf [accessed 2023-07-18]
24. Emery CA. Injury prevention and future research. Med Sport Sci 2005;48:179-200. [doi: 10.1159/000084289] [Medline: 16247258]
25. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Clin J Sport Med 2006 Mar;16(2):97-106. [doi: 10.1097/00042752-200603000-00003] [Medline: 16603877]

Abbreviations

NFL: National Football League

Edited by E Meinert; submitted 12.01.23; peer-reviewed by A Samaranayaka, Anonymous, Anonymous, F Pereira; revised version received 24.07.23; accepted 27.12.23; published 13.02.24.

Please cite as:
Puga TB, Schafer J, Thiel G, Scigliano N, Ruan T, Toledo A, Agbedanu PN, Treffer K
COVID-19 National Football League (NFL) Injury Analysis: Follow-Up Study
JMIRx Med 2024;5:e45688
URL: https://xmed.jmir.org/2024/1/e45688
doi:10.2196/45688

The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study

Silvia Corbera[*], PhD; Amanda M Marín-Chollom[*], PhD

Department of Psychological Science, Central Connecticut State University, New Britain, CT, United States

[*]all authors contributed equally

Corresponding Author:
Silvia Corbera, PhD

Related Articles:

http://preprints.jmir.org/preprint/52970

https://med.jmirx.org/2024/1/e59430

https://med.jmirx.org/2024/1/e58859

Abstract

Background: In situations of acute stress, individuals may engage in prosocial behaviors or risk-taking self-oriented behaviors. The COVID-19 pandemic created large stress-promoting conditions that impacted individuals' decisions to adhere to COVID-19 preventative behaviors.

Objectives: The study aimed to examine the relationship between anxiety during the pandemic and adherence behaviors to prevent the spread of COVID-19, and the moderating influence of prosocial behaviors. We hypothesized that individuals with high anxiety during COVID-19 would adhere more to preventive COVID-19 behaviors than ones with low anxiety and that this relationship would be stronger in those individuals with higher prosocial behaviors.

Methods: A web-based survey was administered through the SONA web-based participant tool of the psychology department of a university in the Northeastern United States. A final sample of 54 undergraduate students completed web-based questionnaires during the second wave of the COVID-19 pandemic, from January to May 2021, which included demographic measures and surveys on prosocial behaviors, anxiety, and COVID-19 preventive behaviors. Moderation analyses were conducted using PROCESS in SPSS.

Results: Participants reported high levels of trait and state anxiety symptoms, most of them meeting or exceeding the cutoff criteria to be clinically meaningful (state anxiety: 47/54, 87%; trait anxiety: 38/44, 86%), and over 50% highly adhered to the COVID-19 preventive behaviors of wearing a face mask, using hand sanitizer, handwashing, coughing/sneezing into their elbow or a tissue, self-quarantining, maintaining social distance, avoiding social gatherings, and avoiding nonessential travel. No significant associations were observed between prosocial behavior, anxiety types, and adherence to COVID-19 preventive behaviors. However, when moderation analyses were conducted between anxiety types and adherence to COVID-19 preventive behaviors, results demonstrated a statistically significant interaction of public prosocial behavior with state anxiety (β=−.17, t_{53}=−2.60; P=.01), predicting engagement in COVID-19 preventative behaviors. At high levels of anxiety, low levels of prosocial public behaviors were associated with higher engagement in COVID-19 preventative behaviors. In contrast, high levels of public prosocial behavior were associated with low engagement in COVID-19 preventative behaviors at higher levels of anxiety.

Conclusions: These results provide information that can aid in the creation of interventions that could increase adherence to COVID-19 preventative behaviors (Reviewed by the Plan P #PeerRef Community).

(JMIRx Med 2024;5:e52970) doi:10.2196/52970

KEYWORDS

prosocial behavior; COVID-19; anxiety; COVID-19 prevention; preventive health behavior; adherence to prevention

Introduction

The understanding of human behavior under stress and in situations of anxiety has been an extensive topic of research [1,2]. Ample evidence suggests that acute stress affects general cognition, executive functioning, working memory [3-5], social-emotional information processing, and social and prosocial behavior [6-8]. Stress also affects decision-making skills [1,2,9-11]; however, this relationship depends on several factors, such as the type of decision-making situation [9], time pressure [11], and the individual characteristics of the participants, such as gender [2,12].

Evidence has also shown that social behavior and social decision-making can be influenced by stress; however, studies are conflicted in this area. Some studies have shown that in situations of acute stress, individuals might engage in more prosocial and empathetic behaviors such as prosocial trustworthiness and sharing [4,7,13,14]; however, on the other hand, other studies suggest that in situations of stress, individuals might engage in unfavorable and risk-taking self-oriented behaviors and distrust [15,16]. Most of the research on the effects of stress in decision-making has been conducted in laboratory studies using stress-inducing paradigms, in which the stress induced was either physical (eg, cold pressure test in which participants immerse their hand in a bucket of cold water) or psychosocial (eg, Trier Social Stress Test in which participants perform a mock job interview). However, to our knowledge, just one study has investigated decision-making in the context of the COVID-19 pandemic as the stressor [17]. Romero-Rivas and Rodriguez-Cuadrado [17] examined whether the psychological impact generated by the COVID-19 pandemic influenced decision-making processes by presenting participants with four decision-making tasks (the dictator game, framing problems, utilitarian/deontological moral dilemmas, and altruistic/egoistic moral dilemmas). This study showed that higher levels of psychological impact were related to safer responses in framing problems and more deontological/altruistic responses to moral dilemmas [17], suggesting that the psychological impact of COVID-19 affected decision-making processes by participants showing safer and more altruistic responses using laboratory decision-making tasks. Given the influence of stress in the decision-making process [1,2], it is imperative to understand this relationship further, especially during the COVID-19 pandemic as it can aid in understanding the factors that influence compliance and adherence to behaviors to prevent the spread of COVID-19 and its variants.

The worldwide COVID-19 pandemic placed humanity in a devastating global health challenge that required the population to adapt to a rapidly changing situation. Gruber and colleagues [18] conceptualized the COVID-19 pandemic as a multidimensional complex stressor that affected both the individual and family, and multiple societal layers, with toxic social stressors such as social isolation and financial loss [18]. Stress research has shown that situations that are uncontrollable and uncertain, and have a social evaluative threat elicit high stress responses [19]. The COVID-19 pandemic created large stress-promoting conditions, such as uncertainty, life threat, loss, and long exposure to anxiety-inducing information [18], generating extensive detrimental psychological and mental health consequences [18,20]. A recent meta-analysis by Luo and colleagues [20] reported a prevalence of anxiety of 33% and depression of 28% of the population during the COVID-19 pandemic, and a study by Ebrahimi and colleagues [21] reported that the prevalence of clinical anxiety increased 2-3 times during the COVID-19 pandemic in comparison to estimations of similar samples before the pandemic. In addition, the COVID-19 pandemic has increased the prevalence of anxiety among undergraduate college students [22,23], especially among the lower socioeconomic groups [22].

To prevent the spread of COVID-19, public health measures were introduced that involved engagement in new behaviors, and the population was expected to adhere to these preventive behaviors (eg, limiting social distance and using masks). Decisions made under different levels of risk and ambiguity can be affected differently by stress [9,24], and during the stress-promoting conditions inherent in the COVID-19 pandemic, the population was confronted with making high-risk decisions about their behaviors, especially decisions related to preventing the spread of COVID-19.

A few studies have examined the predictors of adherence to preventive behaviors to spread COVID-19 [21,25-27]. Recent research by Pollak and colleagues [25] showed that predictors for nonadherence were high levels of current distress and risk factors such as male sex, not having children, attention-deficit/hyperactivity disorder symptoms, smoking, and high levels of risk-taking behavior. Similarly, Ebrahimi and colleagues [21] reported that female sex, older age, worry about significant others, mandatory adherence, and altruistic attitude were associated with higher adherence, and current employment was associated with lower adherence. Besides, altruistic attitudes, empathy, fairness, and gratitude have also been associated with higher levels of adherence to COVID-19 preventative behaviors [21,26,27].

To our knowledge, no studies have examined the relationship between anxiety and adherence to behaviors to prevent the spread of COVID-19 with the moderating role of prosocial tendencies. Therefore, this study aims to investigate the relationship between anxiety during the COVID-19 pandemic and adherence behaviors to prevent the spread of COVID-19, and the moderating influence of prosocial tendencies on this association.

It is hypothesized that individuals with high anxiety during the COVID-19 pandemic are more likely to adhere to the behaviors to prevent the spread of COVID-19 than individuals with low anxiety, and this relationship will be stronger for individuals with high prosocial behavior tendencies than those with low prosocial behavior tendencies.

Methods

Participants

Participants were undergraduate students recruited at a university in the Northeastern United States through the psychology department's web-based subject pool that is managed by SONA. Students from different majors taking psychology courses were

eligible to sign up to participate in this study and received credit in their courses for participation. Data was collected on the web via Select Survey in the Spring 2021 semester during the COVID-19 pandemic from January 2021 to May 2021. In addition to the measures described below, basic demographics of age, race/ethnicity, class standing, gender, and employment status were also collected.

Ethical Considerations

This study was approved by Central Connecticut State University's Institutional Review Board (IRB#20089), and web-based consent was obtained from all participants. Participants were presented with a web-based informed consent, and they acknowledged it by pressing a button to continue the study and before any data collection. Participants were informed that their information would be kept confidential, and their data would be anonymous. Participants were informed that their data would be part of scientific publications. They were granted 1 SONA credit for 15-30 minutes of their time to complete the questionnaire.

Measures

Prosocial Behaviors

The 23-item Prosocial Tendencies Measure [28,29] was developed to be used with college-age students and young adults, and was used to measure six types of prosocial behaviors: compliant (2 items), dire (3 items), altruistic (5 items), public (4 items), emotional (4 items), and anonymous (5 items). Compliant is when prosocial behaviors are done because others asked for help (eg, "I never hesitate to help others when they ask for it"). Dire is when prosocial behaviors are done to help others in emergencies or crises (eg, "I tend to help people who hurt themselves badly"). Altruistic is helping others and expecting little to nothing in return (eg, "I often help even if I don't think I will get anything out of helping"). Public is when prosocial behaviors are completed because others are watching with the motivation to receive external approval (eg, "Helping others when I am in the spotlight is when I work best"). Emotional is when individuals help others because they are in a highly emotional state (eg, "It is most fulfilling to me when I can comfort someone who is very distressed"). Anonymous is when prosocial social behaviors are enacted intentionally without other people's knowledge (eg, "I tend to help needy others most when they do not know who helped them"). Items were answered on a 5-point Likert Scale from 1 (does not describe me at all) to 5 (describes me greatly). Six total sum subscales were calculated and used in the analysis. Higher scores indicate greater prosocial behavior for each subscale. This scale has shown acceptable to good reliability (public subscale: Cronbach α=0.80; anonymous subscale: Cronbach α=0.88; dire subscale: Cronbach α=0.54; emotional subscale: Cronbach α=0.77; compliant subscale: Cronbach α=0.87; altruism subscale: Cronbach α=0.62) and high test-retest reliability with 2-week reliability correlations (coefficients of public: 0.61; anonymous: 0.75; dire: 0.72; emotional: 0.80; compliant: 0.73; altruism: 0.60; all P<.001) [28].

Anxiety

The State-Trait Anxiety Inventory was used to measure both trait and state anxiety [30,31]. The 40-item inventory uses 20 items for trait anxiety (eg, "I am a steady person") and 20 for state anxiety (eg, "I feel at ease"). Of the 40 items, 19 are reverse scored. All items in this study were rated on a 4-point Likert scale from not at all to very much so. Two total sum subscales were calculated, and higher scores indicate greater anxiety for both subscales. Internal consistency coefficients range from 0.86 to 0.95 and test-retest from 0.65 to 0.75 for 2 months [30].

COVID-19 Preventative Behaviors

The COVID-19 International Survey from the PhenX Toolkit [32,33] was used to specifically collect data on what preventative COVID-19 behaviors participants were engaging in. For this study, we selected 23 items within the survey specific to what frequency individuals were engaging in COVID-19 preventative behaviors (eg, hand washing, mask wearing, physical distancing, avoiding social gatherings, self-quarantining after travel, or self-quarantining if infected or likely infected). These items are answered on a 4-point Likert scale from never to most of the time, with the additional option of "don't know/I prefer not to answer/Not applicable." One total sum score was calculated, and higher scores indicated higher engagement in COVID-19 preventative behaviors.

Analysis

Bivariate correlations between the main variables of interest were first examined. Moderation analyses were then conducted using PROCESS version 3.5 in SPSS Version 29 (IBM Corp). Bootstrapping was set to 5000 and the CI to 95%. Mean-deviated predictor variables (ie, prosocial behaviors and anxiety types) were created and used for all moderation analyses. A total of 12 separate models were tested. One for each type of the six prosocial behaviors, interacting with the two subtypes of anxiety measured that predict engagement in COVID-19–related preventative behaviors. All significant interactions were plotted, using 1 SD below and above the mean for prosocial behaviors, for interpretation. The parameters for using moderation in PROCESS are like any moderation analysis program [34].

Results

Participants were (n=54) college students (mean age 20.74, SD 5.16, range 18-54 years; 1 student was 54 years of age). Most students' ages were closer to the mean (25% percentile: n=19; percentile 50%: n=19; percentile 75%: n=21). Students were mostly enrolled full-time at the university (n=43); employed (n=43); and living at home with their parent, relative, or guardian (n=40). Most of the sample self-identified as women (n=43; men: n=9; transgender: n=1) and White (n=34; Hispanic/Latino/a: n=11; Black/African American: n=7; Asian: n=2). See Table 1 for more demographic information.

Table . Demographic characteristics.

Variable		Participants (N=54), n (%)
Gender		
	Women	43 (80)
	Men	9 (17)
	Transgender	1 (2)
Race/ethnicity		
	White	34 (63)
	Hispanic/Latino/a	11 (20)
	Black/African American	7 (13)
	Asian	2 (4)
Enrollment status		
	Full-time	43 (80)
	Part-time	10 (20)
First generation college student		
	Yes	22 (41)
	No	32 (59)
Marital status		
	Single	40 (74)
	In a relationship	13 (24)
	Married	1 (2)
Employed		
	Yes	43 (80)
	No	10 (20)
Hours worked per week		
	<5	4 (9)
	5-10	2 (4)
	10-20	17 (37)
	20-30	17 (37)
	>30	6 (13)
Housing		
	On campus (residence hall)	8 (15)
	Off-campus housing (within 5 miles of campus)	6 (11)
	Off-campus housing (farther than 5 miles from campus)	39 (72)
Living situation		
	Living alone	5 (9)
	Living with students	5 (9)
	Living with parents/guardians/relatives	40 (74)
	Living with spouse	4 (7)

Participants on average reported high levels of anxiety symptoms on both the trait (mean 51.05, SD 10.37) and state subscales (mean 50.63, SD 11.11). Using the recommended cutoff of 40 to determine clinically meaningful anxiety symptomology, most of the sample either met or exceeded the criteria (state anxiety: 47/54, 87%; trait anxiety: 38/44, 86%). On average, participants highly adhered to COVID-19 preventative actions and behaviors (mean 60.28, SD 13.29). Over 50% of the sample (N=54) engaged in the following behaviors “most of the time”: wearing a face mask (n=49, 91%),

using hand sanitizer (n=45, 83%), handwashing with soap and water (n=44, 82%), coughing/sneezing into their elbow (n=44, 82%), self-quarantining if they have or believe they have the virus (n=42, 78%), self-quarantining if they are returning from a trip (n=34, 63%), coughing/sneezing into a tissue and throwing it away and washing hands (n=37, 69%), staying 6 feet apart from other people (n=29, 54%), avoiding large social gatherings (n=29, 54%), and avoiding any nonessential travel (n=29, 54%; Table 2).

Table . Engagement in COVID-19 preventative behaviors in the past 7 days as reported by participants (N=54).

Action or behavior	Participants responding "most of the time," n (%)
Wearing a face mask	49 (91)
Use hand sanitizer	45 (83)
Handwashing with soap and water	44 (82)
Coughing/sneezing into your elbow	44 (82)
Self-quarantining if you have or believe you have the virus	42 (78)
Coughing/sneezing into a tissue throwing it away and wash hands	37 (69)
Self-quarantining if you are returning from a trip	34 (63)
Staying 6 feet apart from other people	29 (54)
Avoid large social gatherings	29 (54)
Avoiding any nonessential travel	29 (54)
Avoiding going out to bars/pubs	26 (48)
Avoiding using public transportation	24 (44)
Staying/working at home	21 (39)
Avoiding playdates	16 (30)
Exercise outside alone or with people you live with only	16 (30)
Avoiding going to restaurants	12 (22)
Avoiding taking your children to the park	11 (20)
Avoiding all social gatherings (large and small)	10 (19)
Wearing gloves every time you go out of your home	6 (11)
Avoiding going to the grocery store or pharmacy	6 (11)
Avoiding opening the mail or delivered goods	6 (11)
Avoiding getting take-out food or delivery	5 (9)
Avoiding going for walks	1 (2)

None of the prosocial behavior types or anxiety types were associated with engagement in COVID-19 preventative behaviors (see Table 3 for descriptive and correlations of study variables). However, in the one model tested with public prosocial behavior, the interaction of public prosocial behavior with state anxiety predicting engagement in COVID-19 preventative behaviors was statistically significant (β=−.17; *P*=.01), which suggested a crossover effect (Figure 1).

Table . Descriptive statistics and correlations for study variables.

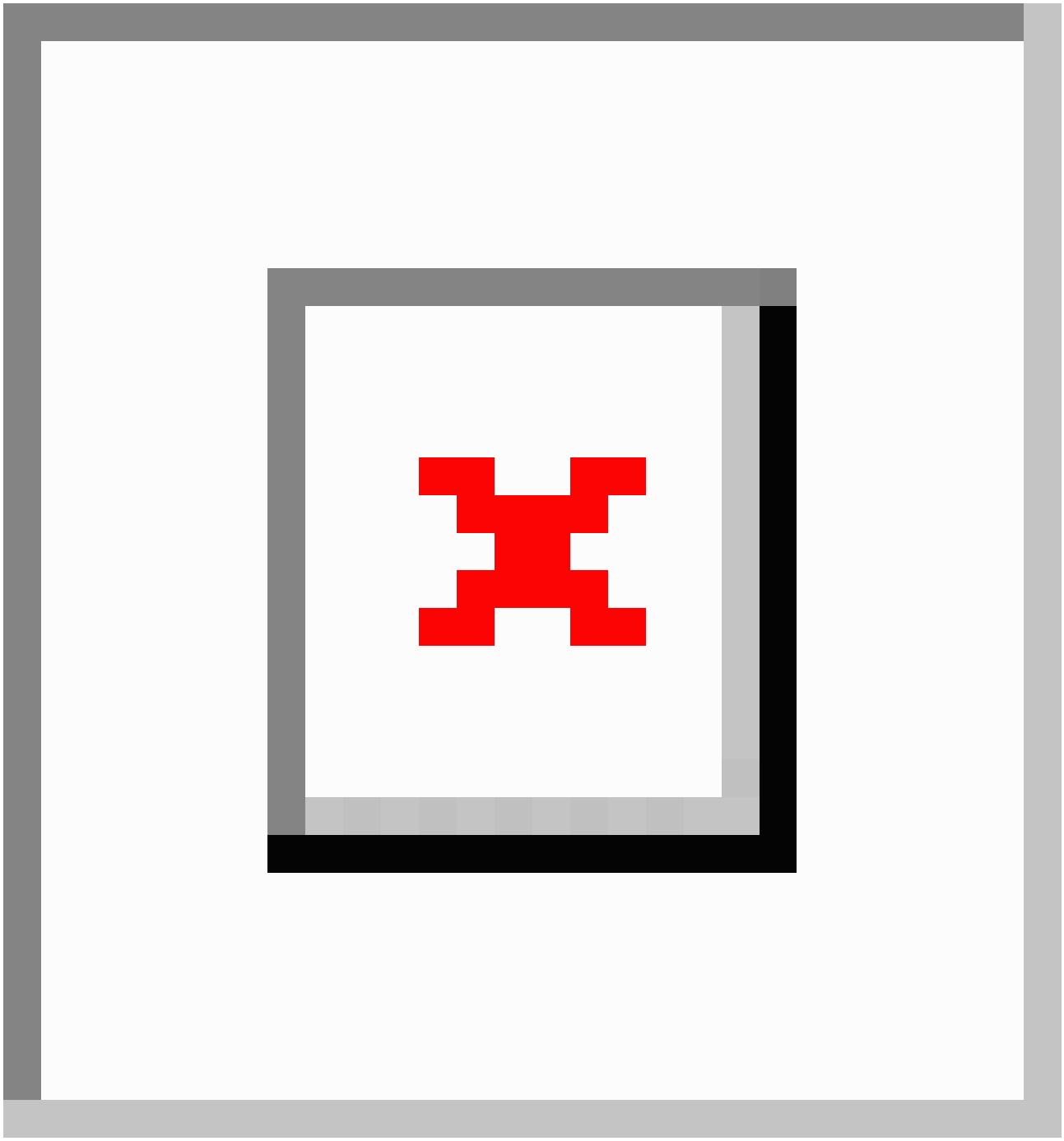
Variable	Partici-pants, n	Mean (SD)	1	2	3	4	5	6	7	8	9
1. COVID-19 PB [a]	54	60.28 (13.29)									
r			—[b]	0.06	0.01	–0.16	0.04	–0.03	0.16	–0.03	–0.01
P value			—	.66	.93	.26	.78	.85	.26	.87	.97
2. Public PT [c]	54	8.26 (3.48)									
r			0.06	—	–0.11	–0.78	–0.05	–0.26	0.26	–0.03	0.05
P value			.66	—	.41	<.001	.75	.05	.06	.84	.74
3. Emotional PT	54	13.93 (3.33)									
r			0.01	–0.11	—	0.09	0.64	0.64	0.49	–0.16	–0.21
P value			.93	.41	—	.53	<.001	<.001	<.001	.31	.12
4. Altruistic PT	54	19.83 (4.15)									
r			–0.16	–0.78	0.09	—	0.02	0.31	–0.20	–0.06	–0.09
P value			.26	<.001	.53	—	.86	.02	.14	.72	.53
5. Dire PT	54	10.35 (2.63)									
r			0.04	–0.05	0.64	0.02	—	0.64	0.55	0.03	0.11
P value			.78	.75	<.001	.86	—	<.001	<.001	.84	.43
6. Compliant PT	54	7.89 (1.76)									
r			–0.03	–0.26	0.64	0.31	0.64	—	0.38	–0.23	–0.20
P value			.85	.05	<.001	.02	<.001	—	.005	.14	.14
7. Anonymous PT	54	14.74 (4.34)									
r			0.16	0.26	0.49	–0.20	0.55	0.38	—	–0.13	–0.18
P value			.26	.06	<.001	.14	<.001	.005	—	.42	.19
8. Trait anxiety	44	51.05 (10.37)									
r			–0.03	–0.03	–0.16	–0.06	0.03	–0.23	–0.13	—	0.84
P value			.87	.84	.31	.72	.84	.14	.42	—	<.001
9. State anxiety	54	50.63 (11.11)									
r			–0.01	0.05	–0.21	–0.09	0.11	–0.20	–0.18	0.84	—
P value			.97	.74	.12	.53	.43	.14	.19	<.001	—

[a]PB: preventative behaviors.

[b]Not applicable.

[c]PT: prosocial tendency.

Figure 1. Public prosocial behavior moderating the association between state anxiety and engagement in COVID-19 preventative behaviors.



At low levels of prosocial public behaviors, engagement in COVID-19 preventative behavior increases as anxiety increases; in contrast, at high levels of prosocial public behaviors, engagement in COVID-19 preventative behavior decreases as anxiety increases. In the other 11 models tested, none of the interaction effects with prosocial behaviors and anxiety-predicting adherence to COVID-19 preventative behavior were statistically significant. (Tables 4 and 5).

Table . Association between state anxiety and COVID-19 preventative behaviors moderated by six types of prosocial behaviors (N=54).[a]

Model		Estimate (SE)	t test (*df*)	*P* value
Complaint		−0.58 (1.12)	−0.52 (53)	.61
	SA[b]	−0.88 (0.75)	−1.18 (53)	.25
	Complaint × SA	0.13 (0.11)	1.19 (53)	.24
Emotional		−0.25 (0.60)	−0.42 (53)	.68
	SA	−0.78 (0.55)	−1.42 (53)	.16
	Emotional × SA	0.07 (0.05)	1.49 (53)	.14
Public		0.41 (0.51)	0.80 (53)	.43
	SA	−0.33 (0.20)	–1.65 (53)	.10
	Public × SA	0.17 (0.06)	−2.60 (53)	.01
Altruistic		−0.60 (0.44)	−1.38 (53)	.17
	SA	−1.76 (0.94)	−1.88 (53)	.07
	Altruistic × SA	−0.10 (0.05)	1.88 (53)	.07
Dire		0.31 (0.73)	0.42 (53)	.68
	SA	−0.33 (0.48)	−0.69 (53)	.49
	Dire × SA	0.04 (0.06)	0.71 (53)	.50
Anonymous		0.51 (0.44)	1.16 (53)	.25
	SA	−0.16 (0.42)	−0.37 (53)	.71
	Anonymous × SA	−0.02 (0.04)	−0.48 (53)	.63

[a]Dependent variable: COVID-19 preventative behaviors. All variables were mean deviated.

[b]SA: state anxiety.

Table . Association between trait anxiety and COVID-19 preventative behaviors moderated by the six types of prosocial behaviors (n=44).[a]

Model		Estimate (SE)	t test (*df*)	*P* value
Complaint		−0.78 (1.24)	−0.63 (43)	.53
	TA[b]	−0.75 (0.86)	−0.88 (43)	.39
	Complaint × TA	0.11 (0.13)	0.83 (43)	.42
Emotional		−0.25 (0.69)	−0.37 (43)	.72
	TA	−0.16 (0.66)	−0.24 (43)	.81
	Emotional × TA	0.01 (0.06)	0.18 (43)	.86
Public		0.35 (0.63)	0.55 (43)	.58
	TA	−0.27 (0.24)	−1.13 (43)	.26
	Public × TA	−0.14 (0.07)	−1.88 (43)	.07
Altruistic		−0.56 (0.53)	−1.06 (43)	.30
	TA	−1.15 (1.11)	–1.04 (43)	.30
	Altruistic × TA	0.06 (0.06)	1.02 (43)	.31
Dire		−0.31 (0.87)	−0.36 (43)	.72
	TA	−0.62 (0.78)	−0.79 (43)	.43
	Dire × TA	0.07 (0.08)	0.78 (43)	.44
Anonymous		0.33 (0.54)	0.61 (43)	.55
	TA	0.18 (0.49)	0.36 (43)	.72
	Anonymous × TA	−0.02 (0.04)	−0.44 (43)	.66

[a]Dependent variable: COVID-19 preventative behaviors. All variables were mean deviated, except the dependent variable.

[b]TA: trait anxiety.

Discussion

Principal Findings

This study aimed to examine the association between state and trait anxiety during the COVID-19 pandemic and the adherence behaviors to prevent the spread of COVID-19 (eg, handwashing and mask wearing), and to investigate the moderating role of prosocial behaviors on this association. Results revealed a statistically significant interaction of public prosocial behaviors with state anxiety predicting adherence to COVID-19 preventative behaviors. This interaction showed that only at low levels of prosocial behaviors (ie, lower self-oriented tendencies with less approval-seeking tendencies), as anxiety increased, engagement in COVID-19 preventative behavior increased, and in contrast, at high levels of prosocial public behaviors (ie, higher self-oriented tendencies with more approval-seeking tendencies), as anxiety increased, adherence to COVID-19 preventative behavior decreased. More specifically, our results revealed that there were no associations between state and trait anxiety and any of the prosocial behaviors (emotional, altruism, dire, anonymous, public, and compliant), and only public prosocial tendencies showed a significant interaction with state anxiety, predicting engagement in COVID-19 preventative behaviors, and suggesting a crossover effect.

Comparison With Prior Work

Based on previous literature on decision-making under stress-promoting conditions [4,7,13,14], we hypothesized that participants with high state and trait anxiety would be more likely to adhere to the behaviors to prevent the spread of COVID-19 than those with low state and trait anxiety, and that this relationship would be stronger for individuals with high prosocial behavior tendencies than those with low prosocial tendencies. Because of the exceptional situation created by the COVID-19 pandemic, we did not specifically predict that any type of prosocial behavior would show a stronger relationship than another.

Concerning the lack of association between state and trait anxiety and any of the prosocial behaviors, these results were not surprising, as we expected that differential levels of prosocial tendencies would relate differently with anxiety when predicting adherence behaviors and therefore that they would operate as a moderator. On the other hand, the interaction that partially supported our hypothesis was with public prosocial behaviors, which moderated the relationship between state anxiety and adherence behaviors to prevent COVID-19. This moderation effect demonstrated a crossover effect. More specifically, at low levels of prosocial public behaviors (ie, lower self-oriented tendencies with less approval-seeking tendencies), as anxiety increases, engagement in COVID-19 preventative behavior increases; in contrast, at high levels of prosocial public behaviors (ie, higher self-oriented tendencies with more approval-seeking tendencies), as anxiety increases, engagement in COVID-19

preventative behavior decreases. As described in the validation study of the Prosocial Tendencies Measure of Carlo and Randall [28], the public prosocial scale was significantly negatively correlated with the altruism, anonymous, and the compliant prosocial subscales. Additionally, public prosocial behaviors were shown to be inversely associated with sympathy, ascription of responsibility, perspective taking, and internalized prosocial moral reasoning, and on the other hand, positively associated with approval-oriented prosocial moral reasoning and hedonism [28]. Therefore, individuals endorsing higher public prosocial tendencies would show more approval-seeking behaviors and be more oriented toward their own needs (as opposed to other people's needs).

Even though we did not find a direct relationship between altruistic, prosocial, or compliant prosocial behaviors, which are the types that capture other-oriented, sympathetic, and morally oriented behaviors, and anxiety and adherence to COVID-19 preventive behaviors, the results revealed that individuals with low levels of public and other approval-seeking behaviors would be more likely to engage in COVID-19 preventive behaviors in situations with high anxiety, which partially supports our hypothesis. The lack of effects with trait anxiety could be due to the smaller sample size of participants answering the trait anxiety scale (n=44) compared to the larger size (N=54) that completed the state anxiety scale, or state anxiety may be more relevant to the COVID-19 pandemic context as it was situational. Even though both types of anxiety were high in this sample (87% of the sample in state anxiety and 86% of the sample in trait anxiety were above the cutoff of 40), state anxiety may have a bigger influence on prosocial and adherence behaviors within this context.

In addition, given the inverse relationship between public and altruistic prosocial behaviors [28], these results suggest that individuals low on public prosocial behaviors would show more other-oriented behaviors. On the other hand, our study did not find any differential relationship between individuals with altruistic prosocial behaviors and state anxiety and adherence behaviors, and therefore, more research is needed in this area. Our results are in partial agreement with the study of Romero-Rivas and Rodriguez-Cuadrado [17], which found that high states of psychological impact from the COVID-19 pandemic were related to individuals being more risk averse and more altruistic in decision-making. Our results showed higher state anxiety was related to more engagement in adherence behaviors to prevent COVID-19 in only the individuals with lower public prosocial behaviors (more other-oriented and less approval-seeking behaviors). Therefore, we propose that higher anxiety relates to prosocial decision-making differently depending on the individual characteristics, in which the individuals with more other-oriented and less approval-seeking prosocial tendencies with higher state anxiety would be more likely to adhere to COVID-19 preventive behaviors. On the other hand, individuals with higher levels of self-oriented or approval-seeking behaviors (public prosocial) show less adherence to COVID-19 preventive behaviors with an increase in state anxiety. As the group with high public prosocial tendencies has been reported to be motivated by self-oriented and approval-seeking tendencies [28], we speculate that situations with high state anxiety would prevent these individuals from engaging in behaviors that would necessitate other-oriented prosocial behaviors and, therefore, would prevent them from engaging in prosocial decision-making and other-oriented behaviors.

Strengths and Future Directions

We suggest that more studies are needed to explore the role of each type of prosocial behavior tendency in addition to a global measure of prosocial tendencies in the engagement of adherence behaviors to prevent COVID-19. Additionally, other variables that may have influenced the results such as vaccine hesitancy, perceived personal risk and disease vulnerability, and trust in science may be potential variables to study in future research, especially regarding the factors that may impact the adherence to preventive behaviors in young adults [35,36]. Our results further the current knowledge on the factors that predict the engagement to adherence behaviors to prevent COVID-19 and provide insights on the cultivation and creation of interventions to promote other-oriented versus self-oriented motivations and tendencies, and in the creation of anxiety-reducing interventions that could increase adherence to COVID-19 preventative behaviors.

Limitations

Our study has several limitations. First, the sample size was small, with 54 participants. However, the performed power analysis, with a power (β) set at 0.8, the Cohen *f* at 0.15 for medium effects, and significance level at .05 (SPSS v29), indicated that the sample needed was 55. Given that the study was conducted during the second wave of the COVID-19 pandemic, we obtained these participants at a critical time point, and we did not recruit additional participants to increase the study sample as the world and life circumstances changed substantially after spring 2021. Second, the sampling method used was convenience sampling with an undergraduate college population, which could have led to the collection of data from participants who were more motivated and self-selected. Additionally, this convenience sampling method led to the recruitment of a homogeneous sample comprised of undergraduate students. This may have impacted the generalizability of the results, and we raise caution when extrapolating results across age groups, educational backgrounds, and cultural contexts. Third, given the cross-sectional design of the study and use of web-based self-reports, we cannot assert causal relationships between the variables of the study, which limits the interpretation of the data. Fourth, because of the small sample size and the cross-sectional nature of the study, these results on their own cannot make any definitive public health policy recommendations but can inform other studies on the role of prosocial tendencies in the relationship between stress and adherence to preventive behaviors. Additionally, the study was conducted on the web, and therefore, it was not possible to control for attention and effort in the participants' responses. Finally, the measures were self-report questionnaires that are prone to biases such as social desirability.

Conclusion

To sum up, our results revealed that anxiety was differently associated with prosocial decision-making processes. At low levels of prosocial public behaviors, as state anxiety increased, the engagement in adherence to behaviors to prevent the spread of COVID-19 also increased. On the other hand, at high levels of prosocial public behaviors, as state anxiety increased, the engagement in adherence to behaviors to prevent the spread of COVID-19 decreased. These results provide valuable information that can aid in the creation of interventions to promote other-oriented motivations and in the reduction of anxiety that could increase adherence to COVID-19 preventative behaviors.

Acknowledgments

We want to express our appreciation to our research participants. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The data generated from this study can be available from the corresponding author upon reasonable request.

Editorial Notice

This paper was peer-reviewed by the Plan P Hashtag Community partner #PeerRef.

Authors' Contributions

SC and AMMC contributed substantially to each stage of the research process: project conceptualization and design, and methodology and project administration. SC and AMMC contributed substantially to data curation, data analysis, manuscript drafting, and manuscript reviewing and editing. Generative artificial intelligence was not used in any portion of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Starcke K, Brand M. Decision making under stress: a selective review. Neurosci Biobehav Rev 2012 Apr;36(4):1228-1248. [doi: 10.1016/j.neubiorev.2012.02.003] [Medline: 22342781]
2. Nowacki J, Heekeren HR, Deuter CE, et al. Decision making in response to physiological and combined physiological and psychosocial stress. Behav Neurosci 2019 Feb;133(1):59-67. [doi: 10.1037/bne0000288] [Medline: 30556701]
3. Shields GS, Sazma MA, Yonelinas AP. The effects of acute stress on core executive functions: a meta-analysis and comparison with cortisol. Neurosci Biobehav Rev 2016 Sep;68:651-668. [doi: 10.1016/j.neubiorev.2016.06.038] [Medline: 27371161]
4. Wolf OT, Schulte JM, Drimalla H, Hamacher-Dang TC, Knoch D, Dziobek I. Enhanced emotional empathy after psychosocial stress in young healthy men. Stress 2015;18(6):631-637. [doi: 10.3109/10253890.2015.1078787] [Medline: 26365125]
5. Wingenfeld K, Duesenberg M, Fleischer J, et al. Psychosocial stress differentially affects emotional empathy in women with borderline personality disorder and healthy controls. Acta Psychiatr Scand 2018 Mar;137(3):206-215. [doi: 10.1111/acps.12856] [Medline: 29417987]
6. Frisch JU, Häusser JA, Mojzisch A. The Trier Social Stress Test as a paradigm to study how people respond to threat in social interactions. Front Psychol 2015 Feb 2;6:14. [doi: 10.3389/fpsyg.2015.00014] [Medline: 25698987]
7. von Dawans B, Strojny J, Domes G. The effects of acute stress and stress hormones on social cognition and behavior: current state of research and future directions. Neurosci Biobehav Rev 2021 Feb;121:75-88. [doi: 10.1016/j.neubiorev.2020.11.026] [Medline: 33301780]
8. Gonzalez-Liencres C, Breidenstein A, Wolf OT, Brüne M. Sex-dependent effects of stress on brain correlates to empathy for pain. Int J Psychophysiol 2016 Jul;105:47-56. [doi: 10.1016/j.ijpsycho.2016.04.011] [Medline: 27150848]
9. Starcke K, Brand M. Effects of stress on decisions under uncertainty: a meta-analysis. Psychol Bull 2016 Sep;142(9):909-933. [doi: 10.1037/bul0000060] [Medline: 27213236]
10. de Visser L, van der Knaap LJ, van de Loo AJAE, van der Weerd CMM, Ohl F, van den Bos R. Trait anxiety affects decision-making differently in healthy men and women: towards gender-specific endophenotypes of anxiety. Neuropsychologia 2010 May;48(6):1598-1606. [doi: 10.1016/j.neuropsychologia.2010.01.027] [Medline: 20138896]
11. Soshi T, Nagamine M, Fukuda E, Takeuchi A. Pre-specified anxiety predicts future decision-making performances under different temporally constrained conditions. Front Psychol 2019 Jul 9;10:1544. [doi: 10.3389/fpsyg.2019.01544] [Medline: 31354572]

12. van den Bos R, Harteveld M, Stoop H. Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women. Psychoneuroendocrinology 2009 Nov;34(10):1449-1458. [doi: 10.1016/j.psyneuen.2009.04.016] [Medline: 19497677]
13. von Dawans B, Ditzen B, Trueg A, Fischbacher U, Heinrichs M. Effects of acute stress on social behavior in women. Psychoneuroendocrinology 2019 Jan;99:137-144. [doi: 10.1016/j.psyneuen.2018.08.031] [Medline: 30240980]
14. von Dawans B, Fischbacher U, Kirschbaum C, Fehr E, Heinrichs M. The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. Psychol Sci 2012 Jun;23(6):651-660. [doi: 10.1177/0956797611431576] [Medline: 22593119]
15. Bendahan S, Goette L, Thoresen J, Loued-Khenissi L, Hollis F, Sandi C. Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk. Eur J Neurosci 2017 Apr;45(7):877-885. [doi: 10.1111/ejn.13395] [Medline: 27606489]
16. Steinbeis N, Engert V, Linz R, Singer T. The effects of stress and affiliation on social decision-making: investigating the tend-and-befriend pattern. Psychoneuroendocrinology 2015 Dec;62:138-148. [doi: 10.1016/j.psyneuen.2015.08.003] [Medline: 26311359]
17. Romero-Rivas C, Rodriguez-Cuadrado S. The psychological impact of the COVID-19 pandemic affected decision-making processes. Span J Psychol 2021 Mar 22;24:e16. [doi: 10.1017/SJP.2021.14] [Medline: 33745483]
18. Gruber J, Prinstein MJ, Clark LA, et al. Mental health and clinical psychological science in the time of COVID-19: challenges, opportunities, and a call to action. Am Psychol 2021 Apr;76(3):409-426. [doi: 10.1037/amp0000707] [Medline: 32772538]
19. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004 May;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]
20. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public – a systematic review and meta-analysis. Psychiatry Res 2020 Sep;291:113190. [doi: 10.1016/j.psychres.2020.113190] [Medline: 32563745]
21. Ebrahimi OV, Hoffart A, Johnson SU. Physical distancing and mental health during the COVID-19 pandemic: factors associated with psychological symptoms and adherence to pandemic mitigation strategies. Clin Psychol Sci 2021 Mar 12;9(3):489-506. [doi: 10.1177/2167702621994545]
22. Rudenstine S, McNeal K, Schulder T, et al. Depression and anxiety during the COVID-19 pandemic in an urban, low-income public university sample. J Trauma Stress 2021 Feb;34(1):12-22. [doi: 10.1002/jts.22600] [Medline: 33045107]
23. Lee J, Solomon M, Stead T, Kwon B, Ganti L. Impact of COVID-19 on the mental health of US college students. BMC Psychol 2021 Jun 8;9(1):95. [doi: 10.1186/s40359-021-00598-3] [Medline: 34103081]
24. Brand M, Labudda K, Markowitsch HJ. Neuropsychological correlates of decision-making in ambiguous and risky situations. Neural Netw 2006 Oct;19(8):1266-1276. [doi: 10.1016/j.neunet.2006.03.001] [Medline: 16942857]
25. Pollak Y, Dayan H, Shoham R, Berger I. Predictors of non‐adherence to public health instructions during the COVID‐19 pandemic. Psychiatry Clin Neurosci 2020 Nov;74(11):602-604. [doi: 10.1111/pcn.13122] [Medline: 32729646]
26. Pfattheicher S, Nockur L, Böhm R, Sassenrath C, Petersen MB. The emotional path to action: empathy promotes physical distancing and wearing of face masks during the COVID-19 pandemic. Psychol Sci 2020 Nov;31(11):1363-1373. [doi: 10.1177/0956797620964422] [Medline: 32993455]
27. Syropoulos S, Markowitz EM. Prosocial responses to COVID-19: examining the role of gratitude, fairness and legacy motives. Pers Individ Dif 2021 Mar;171:110488. [doi: 10.1016/j.paid.2020.110488] [Medline: 35502308]
28. Carlo G, Randall BA. The development of a measure of prosocial behaviors for late adolescents. J Youth Adolesc 2002 Feb;31(1):31-44. [doi: 10.1023/A:1014033032440]
29. Carlo G, Knight GP, McGinley M, Zamboanga BL, Jarvis LH. The multidimensionality of prosocial behaviors and evidence of measurement equivalence in Mexican American and European American early adolescents. J Res Adolesc 2010 May 18;20(2):334-358. [doi: 10.1111/j.1532-7795.2010.00637.x]
30. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory: Spielberger; 1983.
31. Spielberger CD. State-Trait Anxiety Inventory: Bibliography: Consulting Psychologists Press; 1989.
32. Hamilton CM, Strader LC, Pratt JG, et al. The PhenX Toolkit: get the most from your measures. Am J Epidemiol 2011 Aug 1;174(3):253-260. [doi: 10.1093/aje/kwr193] [Medline: 21749974]
33. Bacon SL, Lavoie KL, Boyle J, Stojanovic J, Joyal-Desmarais K, iCARE study team. International assessment of the link between COVID-19 related attitudes, concerns and behaviours in relation to public health policies: optimising policy strategies to improve health, economic and quality of life outcomes (the iCARE study). BMJ Open 2021 Mar 11;11(3):e046127. [doi: 10.1136/bmjopen-2020-046127] [Medline: 33707274]
34. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis Third Edition: A Regression-Based Approach: Guilford Press; 2022.
35. Gupta S, Watanabe S, Laurent SM. Psychological predictors of vaccination intentions among U.S. undergraduates and online panel workers during the 2020 COVID-19 pandemic. PLoS One 2021 Nov 30;16(11):e0260380. [doi: 10.1371/journal.pone.0260380] [Medline: 34847162]

36. Hromatko I, Tonković M, Vranic A. Trust in science, perceived vulnerability to disease, and adherence to pharmacological and non-pharmacological COVID-19 recommendations. Front Psychol 2021 Apr 30;12:664554. [doi: 10.3389/fpsyg.2021.664554] [Medline: 33995222]

Edited by T Leung; submitted 20.09.23; peer-reviewed by D Saderi, FQ Arogundade, L Ram, M Tomaskova, M Pulier, S Azra; revised version received 25.03.24; accepted 25.03.24; published 27.05.24.

Please cite as:
Corbera S, Marín-Chollom AM
The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study
JMIRx Med 2024;5:e52970
URL: https://xmed.jmir.org/2024/1/e52970
doi:10.2196/52970

Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study

Tatsuki Onishi[1,2,3], MPH, MD

1

2

3

Corresponding Author:
Tatsuki Onishi, MPH, MD

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.01.13.24301276v1

Companion article: https://med.jmirx.org/2024/1/e63862

Companion article: https://med.jmirx.org/2024/1/e63859

Abstract

Background: Sleeve gastrectomy is an effective surgical option for morbid obesity, and it improves glucose homeostasis. In patients with gastric cancer and type 2 diabetes mellitus (DM), gastrectomy, including total gastrectomy, is beneficial for glycemic control.

Objective: This study aims to clarify the effects of gastrectomy and different reconstructive techniques on the incidence of postoperative DM in patients with gastric cancer.

Methods: This retrospective, single-center, cohort study included 715 patients without DM who underwent total gastrectomy at the Tokyo Metropolitan Bokutoh Hospital between August 2005 and March 2019. Patients underwent reconstruction by Roux-en-Y (RY) gastric bypass or other surgical techniques (OT), with DM onset determined by hemoglobin A_{1c} levels or medical records. Analyses included 2-sample, 2-tailed *t* tests; χ^2 tests; and the Kaplan-Meier method with log-rank tests to compare the onset curves between the RY and OT groups, along with additional curves stratified by sex. A Swimmer plot for censoring and new-onset DM was implemented.

Results: Stratified data analysis compared the RY and OT reconstruction methods. The hazard ratio was 1.52 (95% CI 1.06-2.18; *P*=.02), which indicated a statistically significant difference in the incidence of new-onset diabetes between the RY and OT groups in patients with gastric cancer. The hazard ratio after propensity score matching was 1.42 (95% CI 1.09-1.86; *P*=.009).

Conclusions: This first-of-its-kind study provides insight into how different methods of gastric reconstruction affect postoperative diabetes. The results suggest significant differences in new-onset DM after surgery based on the reconstruction method. This research highlights the need for careful surgical planning to consider potential postoperative DM, particularly in patients with a family history of DM. Future studies should investigate the role of gut microbiota and other reconstructive techniques, such as laparoscopic jejunal interposition, in developing postoperative DM.

(JMIRx Med 2024;5:e56405) doi:10.2196/56405

KEYWORDS

diabetes mellitus; gastrectomy; gastric cancer surgery; glucose metabolism; postoperative diabetes onset; surgery outcomes

Introduction

Gastrectomy, particularly sleeve gastrectomy (SG), has been shown to be an effective surgical option for morbid obesity due to its low complication rates and significant weight loss results [1-5]. SG results in alteration of the appetite through the regulation of gut hormones, resulting in decreased hunger and increased satiety [6]. SG also improves glucose homeostasis through resulting changes in gut hormone levels [7]. Specifically, laparoscopic SG results in significant improvement

in glucose metabolism in patients who are morbidly obese and has been found to stop the development of diabetes at a high rate [8]. SG has been shown to improve blood glucose independently of weight loss by restoring hepatic insulin sensitivity [9]. However, the effects of gastrectomy on patients who are not obese with type 2 diabetes mellitus (DM) are less clear, with some studies suggesting that gastrectomy may improve diabetic status [10].

In patients with gastric cancer diagnosed with type 2 DM, gastrectomy has been found to have a positive impact on their glycemic control. Improvements in glycemic control, or even diabetes remission, have been reported after gastrectomy [10-15]. The extent of the gastrectomy, duration of diabetes, and method of reconstruction have been identified as important factors influencing the improvements in glycemic control [10-14]. Although the mechanisms underlying these effects are not fully understood, oncometabolic surgeries, including gastrectomy, have been suggested as a potential treatment for type 2 DM in patients with gastric cancer [16].

Studies have shown that total gastrectomy (TG) is associated with improved glucose metabolism in patients with gastric cancer, resulting in a lower rate of newly diagnosed diabetes after surgery [17]. However, the effects of gastrectomy on glucose metabolism in patients with and without diabetes have been inconsistent, with some studies reporting significant reductions in fasting blood glucose levels after gastrectomy [18]. Furthermore, SG has been associated with significant reductions in hemoglobin A_{1c} levels in patients without diabetes, suggesting its possible role in the prevention of type 2 DM [19].

In terms of reconstruction after partial gastrectomy in patients with gastric cancer, both Roux-en-Y (RY) and Billroth II reconstructions have been considered acceptable options [20]. RY reconstruction is often preferred for patients with gastric cancer, given that this procedure can lead to decreased reflux gastritis and esophagitis, decreased probability of cancer recurrence, and decreased incidence of surgical complications [21]. RY reconstruction has also been found to be as effective as other methods with respect to nutritional status and postoperative outcomes [22]. In comparison to Billroth II reconstruction, RY has been shown to have similar postoperative complications and better long-term outcomes [23]. Furthermore, RY reconstruction without cutting has been the preferred method in cases of gastritis, bile reflux, and gastric residuals [24].

Various studies have examined the impact of different reconstructive procedures on postoperative complications in patients with gastric cancer. It has been found that long-limb RY bypass reconstruction could lead to improved glycemic control [25], and it has been observed that preexisting DM is associated with postoperative complications [10,26]. Several studies further support the benefits of RY reconstruction, with some indicating it to be more effective than Billroth II reconstruction [27,28]. Additionally, significant improvements in DM control have been associated with RY reconstruction [25,28,29].

Given these, the aim of this study was to investigate the incidence of new-onset DM in patients with gastric cancer after surgery and how this incidence varies with different types of surgical reconstruction, namely, the RY procedure and other alternative reconstruction techniques. While studies have investigated how surgical treatment for gastric cancer affects existing DM [25,28,29], none have investigated the development of new-onset DM in patients without DM; to the best of our knowledge, this is the very first study to do so. Findings from this study could contribute valuable insights into the postoperative outcomes associated with different gastric reconstruction techniques. Such insights are vital for guiding clinical decisions and optimizing patient care, particularly in the context of mitigating the risk of developing DM after gastric surgery. Moreover, findings from this study are expected to have significant implications for both clinical practice and future research in the field of gastric surgery and DM prevention.

Methods

Ethical Considerations

The study was approved by the Tokyo Metropolitan Bokutoh Hospital's ethics committee (30-110) and conducted in accordance with the Declaration of Helsinki. Individual informed consent was waived because the data were deidentified and not trackable. No compensation was given.

Study Participants

The study design was a retrospective, single-center, cohort study. A total of 715 patients who underwent TG as the definitive procedure and as a standby procedure at the Tokyo Metropolitan Bokutoh Hospital between August 2005 and March 2019 and were not diagnosed with DM at the time of surgery were included in the study. Whether the patients would undergo reconstruction through RY gastric bypass or other surgical techniques (OT) was chosen based on the preference of the surgeons, and the patients were grouped accordingly. The definite onset of diabetes in the patients was considered based on previous electronic medical records or when their hemoglobin A_{1c} value was equal to or greater than 6.5 based on laboratory testing. The competing outcome was death. After a meticulous data curation process using Python (version 3.10; Python Software Foundation) that corrects for missing values and ensures appropriate data types, we obtained a dataset that was optimized for analysis and free of common data inconsistencies.

Statistical Analysis

Basic statistical measures such as the mean, median, and SD were computed. Two-sample, 2-tailed *t* tests and chi-square tests were used to assess the difference in demographic characteristics between the 2 groups. In addition, the Kaplan-Meier method was used to estimate the onset function delineating the interval between the TG and the subsequent emergence of new-onset diabetes postoperatively and was augmented with log-rank tests to help compare the onset curves between the RY and OT groups, along with additional curvesstratified by sex, and the same analysis was carried out after propensity score matching. A Swimmer plot for censoring and new-onset DM was implemented. The abovementioned analyses were conducted using Python (version 3.10).

Results

The characteristics of the patients included in the study at the time of the surgery are shown in Table 1. Of the 715 patients who had a gastrectomy, 489 (68.4%) underwent RY reconstruction.

Table . Demographics of study population.

Characteristics	Missing data (N=715), n (%)	OT[a] (n=226)	RY[b] (n=489)	*P* value
Cases (N=715), n (%)	—[c]	226 (31.6)	489 (68.4)	—
Age (years), mean (SD)	—	70.0 (10.4)	68.1 (10.6)	.03[d]
Male sex, n (%)	—	139 (61.5)	363 (74.2)	.001[e]
Height (cm), mean (SD)	—	157.9 (10.8)	160.6 (8.6)	.001[d]
Weight (kg), mean (SD)	—	56.9 (14.8)	57.1 (11.1)	.87[d]
BMI (kg/m^2), mean (SD)	—	24.6 (32.2)	22.1 (3.4)	.25[d]
ASA-PS[f], n (%)	—			.03[e]
1		21 (9.3)	27 (5.5)	
2		167 (73.9)	384 (78.5)	
3		34 (15)	77 (15.7)	
4		4 (1.8)	1 (0.2)	
Total intravenous anesthesia, n (%)	—	6 (2.7)	43 (8.8)	.004[e]
Nerve block, n (%)	—	210 (92.9)	462 (94.5)	.52[e]
Bleeding (mL), mean (SD)	—	362.5 (301.7)	582.8 (639.1)	<.001[d]
Blood transfusion (mL), mean (SD)	—	23.8 (111.9)	108.6 (373.3)	<.001[d]
Urine (mL), mean (SD)	—	364.4 (319.7)	383.0 (321.1)	.47[d]
Infusion (mL), mean (SD)	—	2185.4 (697.0)	2545.5 (929.5)	<.001[d]
Operating room time (min), mean (SD)	—	298.4 (74.4)	317.5 (76.8)	.002[d]
Anesthesia time (mL), mean (SD)	—	275.5 (75.0)	295.7 (76.3)	.001[d]
Operation time (mL), mean (SD)	—	225.2 (69.6)	244.5 (71.8)	.001[d]
T[g] (OT: n=126, RY: n=252), n (%)	337 (47.1)			<.001[e]
1		56 (44.4)	64 (25.4)	
2		46 (36.5)	81 (32.1)	
3		20 (15.9)	88 (34.9)	
4		4 (3.2)	17 (6.7)	
0		0 (0)	2 (0.8)	
M[h] (OT: n=126, RY: n=251), n (%)	338 (47.3)			.35[e]
0		123 (97.6)	248 (98.8)	
1		3 (2.4)	2 (0.8)	
3		0 (0)	1 (0.4)	
N[i] (OT: n=125, RY: n=249), n (%)	341 (47.7)			<.001[e]
0		76 (60.8)	94 (37.8)	
1		29 (23.2)	71 (28.5)	
2		19 (15.2)	63 (25.3)	
3		1 (0.8)	21 (8.4)	
D[j] (OT: n=120, RY: n=234), n (%)	361 (50.5)			.002[e]
0		2 (1.7)	5 (2.1)	
1		72 (60)	88 (37.4)	

Characteristics	Missing data (N=715), n (%)	OT[a] (n=226)	RY[b] (n=489)	*P* value
2		46 (38.3)	139 (59.1)	
3		0 (0)	2 (0.9)	

[a]OT: other surgical techniques.

[b]RY: Roux-en-Y reconstruction.

[c]Not applicable.

[d]2-sample, 2-tailed *t* test.

[e]χ^2 test.

[f]ASA-PS: American Society of Anesthesiologists Physical Status.

[g]T: tumor (TNM staging).

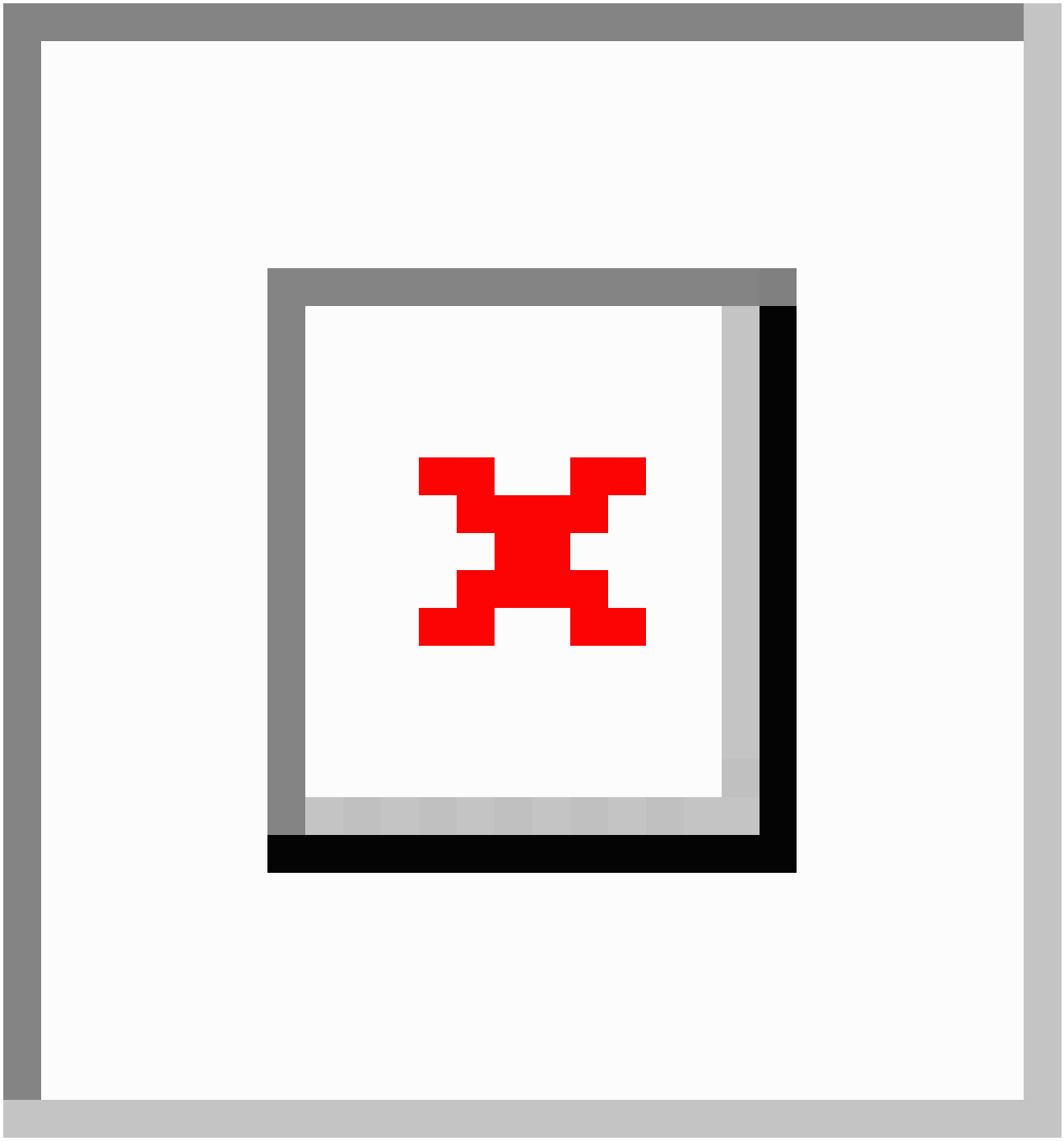
[h]M: metastasis (TNM staging).

[i]N: node (TNM staging).

[j]D: dissection.

The Kaplan-Meier curve of new-onset DM in the RY and OT groups is shown in Figure 1. Granular comparison of the incidence rates of postoperative diabetes associated with these distinct reconstructive procedures was made. The rate of diabetes onset was inferred from the slope of these curves, with a steeper decline indicating a higher incidence within the respective group.

Figure 1. Kaplan-Meier curve of new-onset DM in the RY and OT groups. DM: diabetes mellitus; OT: other surgical techniques; RY: Roux-en-Y.



A log-rank test revealed that the hazard ratio was 1.52 (95% CI 1.06-2.18), and the resultant *P* value from this log-rank test was .02, which denotes a statistically significant difference in the incidence of new-onset diabetes after surgery between patients with gastric cancer who underwent RY reconstruction versus OT. These findings indicate a difference in the incidence of postoperative diabetes based on the type of gastric reconstruction method used (Figure 1).

A Swimmer plot was then produced (Figure 2). In the Swimmer plot, orange lines represent RY cases, blue lines represent OT cases, a cross indicates censoring due to death, and a circle represent censoring due to new-onset DM. The last DM onset in the OT group was at approximately 2400 days, which explains the linear part of the Kaplan-Meier curve.

Propensity score matching was conducted according to the use of laparoscopy, age, sex, and BMI. After propensity score matching, the Kaplan-Meier onset curve showed a hazard ratio of 1.42 (95% CI 1.09-1.86), and the resultant *P* value was .009 (Figure 3). This means that the results are robust even when accounting for unknown confounding and that RY cases have more postoperative DM than OT cases.

A Kaplan-Meier curve stratified by sex was also generated (Figure 4). In this Kaplan-Meier curve, there was no significant difference in the development of postoperative DM between the RY and OT groups for both male and female patients (*P*=.12 and *P*=.24, respectively).

Figure 2. Swimmer plot of new-onset DM and death. DM: diabetes mellitus; OT: other surgical techniques; RY: Roux-en-Y.

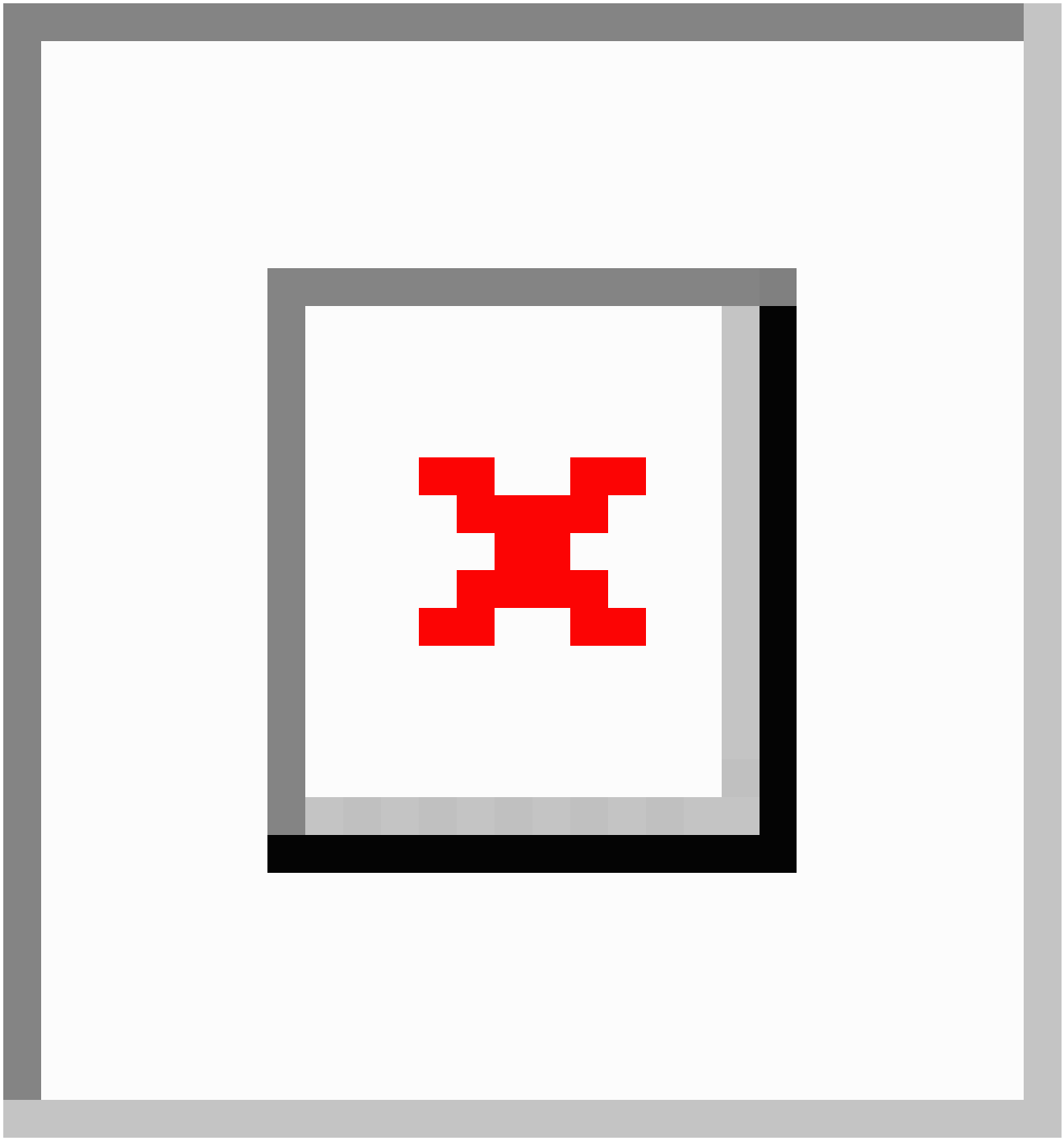


Figure 3. Density histogram (A) before and (B) after propensity score matching, and (C) Kaplan-Meier curve of new-onset DM after propensity score matching. DM: diabetes mellitus; OT: other surgical techniques; RY: Roux-en-Y.

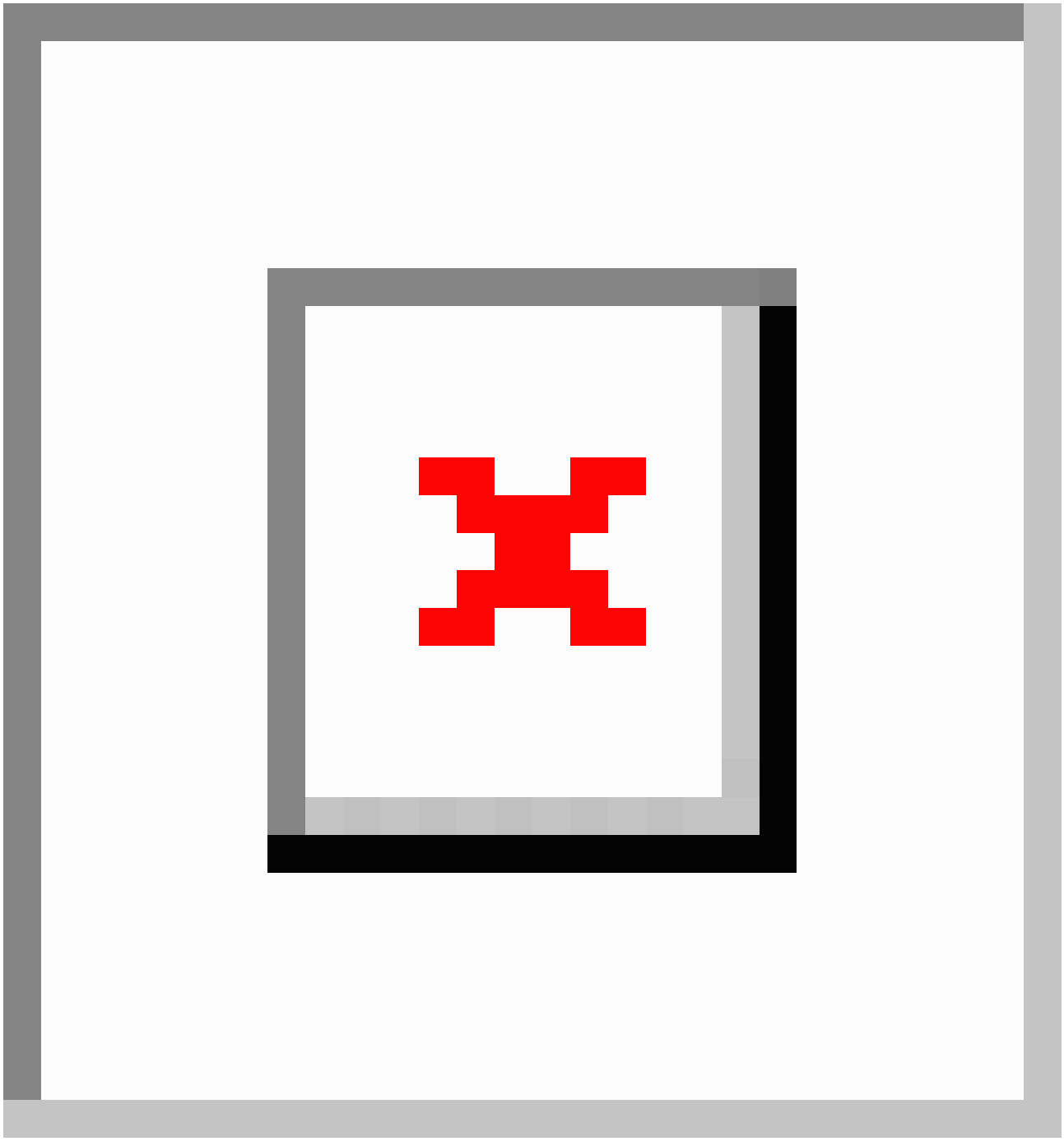
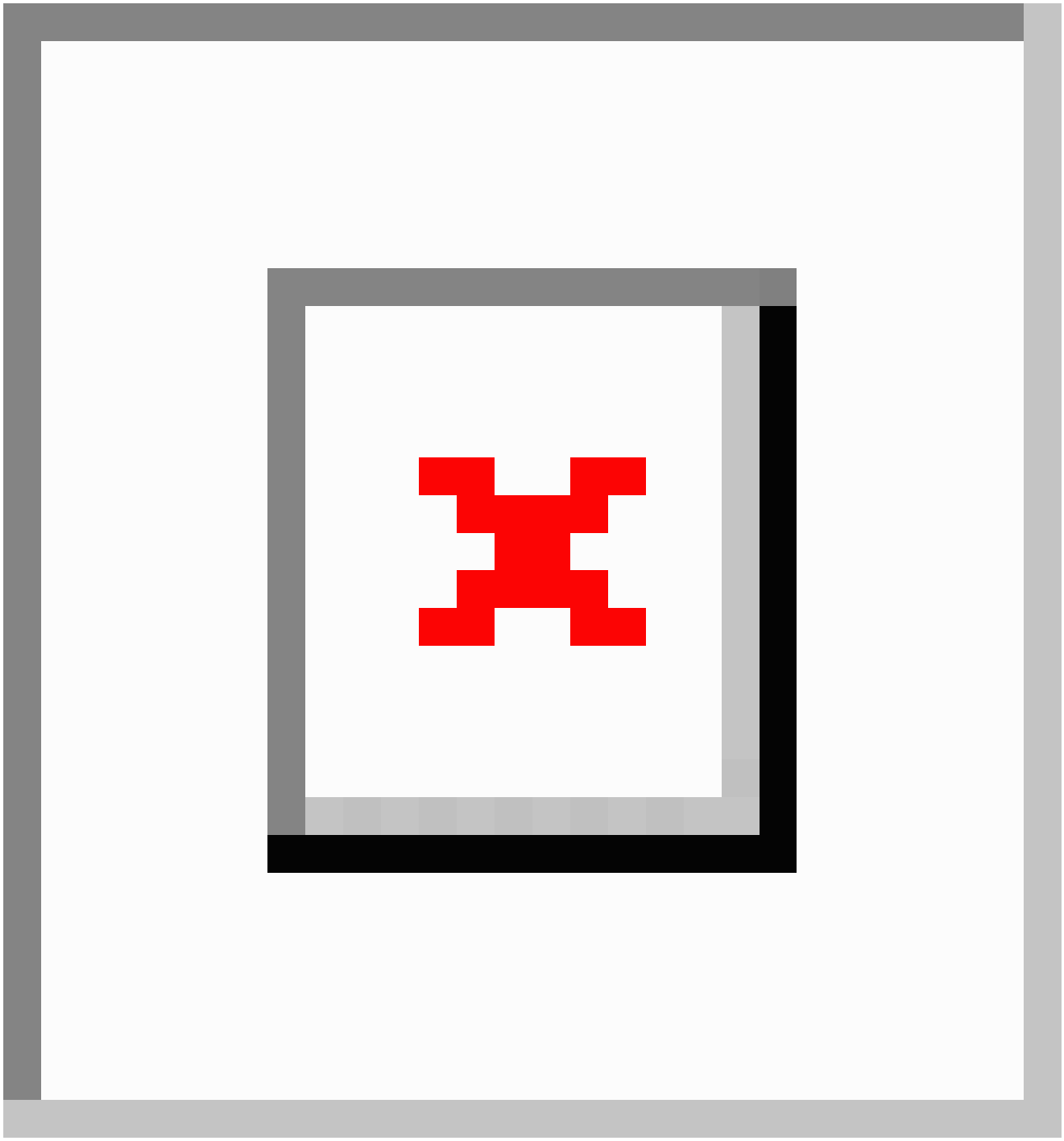


Figure 4. Kaplan-Meier curve of new-onset DM stratified by sex: (A) male and (B) female. DM: diabetes mellitus; OT: other surgical techniques; RY: Roux-en-Y.



Discussion

Principal Findings

This study showed that patients who underwent RY reconstruction had more postoperative DM than those who underwent OT. This study is the first to provide insights into how different methods of gastric reconstruction might affect the risk of developing postoperative DM.

Comparison to Prior Work

Although there were works reporting the influence of preexisting DM after gastrectomy [25,28,29], there was no work regarding new-onset DM after gastrectomy with distinct surgical reconstruction techniques. Some studies mentioned a change in microbiota after gastrectomy [30,31]. Specific changes in the gut microbiota after surgery include increased species richness, decreased butyrate-producing bacteria, and enrichment of certain symbiotic bacteria [32]. The abundance of specific gut bacterial genera has been found to correlate with the population of peripheral immune cells [32-34].

Strengths

A thorough approach to data preprocessing and the use of robust statistical methods will ensure the reliability and validity of these findings in the wider context of gastric surgery and DM research. Our study is important for understanding the temporal dynamics of DM development after gastric surgery and has

significant implications for surgical planning and patient management to prevent postoperative DM, especially in patients with a strong family history of DM.

Limitations

Deaths for any reason was 14.2% (32/226) in the OT group and 21.7% (106/489) in the RY group (P=.02 by χ^2 test), which may have influenced the interpretation of the results. This study did not include an assessment of other determinants that could potentially influence the development of DM, including lifestyle choices and genetic predisposition. It is plausible that there may be a difference in the intrinsic characteristics of DM in patients who present with diabetic symptoms prior to undergoing surgery for gastric cancer, as opposed to those in whom the onset of gastric malignancy precedes the development of DM. Such considerations were beyond the scope of analysis within the parameters of this study.

Future Directions

First, laparoscopic jejunal interposition reconstruction (LJIP), a surgical technique in which a pouch is created in the jejunum to reconstruct the upper gastrointestinal tract, may be appropriate for patients with impaired glucose tolerance [35]. Studies have shown that LJIP leads to better postoperative outcomes, including improved quality of life and nutritional status, compared with other reconstruction methods [35-38]. Future study should include LJIP as a reconstructive method.

Second, it was not possible to study the gut microbiota. With access to a suitable dataset, we would like to investigate the association between gut microbiota and the development of new-onset DM after gastrectomy.

Third, our study could be improved by comparison with a population that has not undergone TG as a control.

Acknowledgments

The author would like to thank Tatsuyoshi Ikenoue at Shiga University, Data Science and AI Innovation Research Promotion Center and Yoshika Onishi at Wellbeing Keiei LLC for proofreading.

Data Availability

Data are available upon request to the corresponding author. Codes have been uploaded to GitHub [39].

Authors' Contributions

TO contributed to conceptualization, writing—original draft, data curation, visualization, and final approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Gumbs AA, Gagner M, Dakin G, Pomp A. Sleeve gastrectomy for morbid obesity. Obes Surg 2007 Jul;17(7):962-969. [doi: 10.1007/s11695-007-9151-x] [Medline: 17894158]
2. Huang R, Ding X, Fu H, Cai Q. Potential mechanisms of sleeve gastrectomy for reducing weight and improving metabolism in patients with obesity. Surg Obes Relat Dis 2019 Oct;15(10):1861-1871. [doi: 10.1016/j.soard.2019.06.022] [Medline: 31375442]
3. Butte JM, Devaud N, Jarufe NP, et al. Sleeve gastrectomy as treatment for severe obesity after orthotopic liver transplantation. Obes Surg 2007 Nov;17(11):1517-1519. [doi: 10.1007/s11695-008-9432-z] [Medline: 18219781]
4. Rawlins L, Rawlins MP, Brown CC, Schumacher DL. Sleeve gastrectomy: 5-year outcomes of a single institution. Surg Obes Relat Dis 2013;9(1):21-25. [doi: 10.1016/j.soard.2012.08.014] [Medline: 23201209]
5. Armstrong J, O'Malley SP. Outcomes of sleeve gastrectomy for morbid obesity: a safe and effective procedure? Int J Surg 2010;8(1):69-71. [doi: 10.1016/j.ijsu.2009.11.004] [Medline: 19944193]
6. Madsbad S, Dirksen C, Holst JJ. Mechanisms of changes in glucose metabolism and bodyweight after bariatric surgery. Lancet Diabetes Endocrinol 2014 Feb;2(2):152-164. [doi: 10.1016/S2213-8587(13)70218-3] [Medline: 24622719]
7. Wang Y, Guo X, Lu X, Mattar S, Kassab G. Mechanisms of weight loss after sleeve gastrectomy and adjustable gastric banding: far more than just restriction. Obesity (Silver Spring) 2019 Nov;27(11):1776-1783. [doi: 10.1002/oby.22623] [Medline: 31545007]
8. Alexandrou A, Athanasiou A, Michalinos A, Felekouras E, Tsigris C, Diamantis T. Laparoscopic sleeve gastrectomy for morbid obesity: 5-year results. Am J Surg 2015 Feb;209(2):230-234. [doi: 10.1016/j.amjsurg.2014.04.006] [Medline: 25034410]
9. Abu-Gazala S, Horwitz E, Ben-Haroush Schyr R, et al. Sleeve gastrectomy improves glycemia independent of weight loss by restoring hepatic insulin sensitivity. Diabetes 2018 Jun;67(6):1079-1085. [doi: 10.2337/db17-1028] [Medline: 29475831]
10. Mirahmadian M, Hasani M, Taheri E, Qorbani M, Hosseini S. Influence of gastric bypass surgery on resting energy expenditure, body composition, physical activity, and thyroid hormones in morbidly obese patients. Diabetes Metab Syndr Obes 2018 Oct 23;11:667-672. [doi: 10.2147/DMSO.S172028] [Medline: 30425544]

11. Radhakrishnan L, Haridas TV. The effect on glycemic status in type 2 diabetes mellitus after gastric cancer surgery. Int Surg J 2021 Jan;8(1):191-195. [doi: 10.18203/2349-2902.isj20205880]
12. An JY, Kim YM, Yun MA, Jeon BH, Noh SH. Improvement of type 2 diabetes mellitus after gastric cancer surgery: short-term outcome analysis after gastrectomy. World J Gastroenterol 2013 Dec 28;19(48):9410-9417. [doi: 10.3748/wjg.v19.i48.9410] [Medline: 24409070]
13. Lee YK, Lee EK, Lee YJ, et al. Metabolic effects of gastrectomy and duodenal bypass in early gastric cancer patients with T2DM: a prospective single-center cohort study. J Clin Med 2021 Sep 4;10(17):4008. [doi: 10.3390/jcm10174008] [Medline: 34501456]
14. Guner A, Cho M, Son T, Kim HI, Noh SH, Hyung WJ. Improved glycemic control with proximal intestinal bypass and weight loss following gastrectomy in non-obese diabetic gastric cancer patients. Oncotarget 2017 Nov 1;8(61):104605-104614. [doi: 10.18632/oncotarget.22262] [Medline: 29262665]
15. Zhang YC, Wei FX, Han W, et al. Impact of sub-gastrectomy on glucose regulation in gastric cancer patients with T2DM: a follow-up study. Int J Diabetes Dev Ctries 2016 Mar;36(1):89-94. [doi: 10.1007/s13410-015-0437-6]
16. Cheng YX, Peng D, Tao W, Zhang W. Effect of oncometabolic surgery on gastric cancer: the remission of hypertension, type 2 diabetes mellitus, and beyond. World J Gastrointest Oncol 2021 Sep 15;13(9):1157-1163. [doi: 10.4251/wjgo.v13.i9.1157] [Medline: 34616520]
17. Bozzetti F, Bonfanti G, Castellani R, et al. Comparing reconstruction with Roux-en-Y to a pouch following total gastrectomy. J Am Coll Surg 1996 Sep;183(3):243-248. [Medline: 8784318]
18. He L, Zhao Y. Is Roux-en-Y or Billroth-II reconstruction the preferred choice for gastric cancer patients undergoing distal gastrectomy when Billroth I reconstruction is not applicable? a meta-analysis. Medicine (Baltimore) 2019 Nov;98(48):e17093. [doi: 10.1097/MD.0000000000017093] [Medline: 31770192]
19. Park JY, Kim YJ. Uncut Roux-en-Y reconstruction after laparoscopic distal gastrectomy can be a favorable method in terms of gastritis, bile reflux, and gastric residue. J Gastric Cancer 2014 Dec;14(4):229-237. [doi: 10.5230/jgc.2014.14.4.229] [Medline: 25580354]
20. Ho TW, Wu JM, Yang CY, Lai HS, Lai F, Tien YW. Total gastrectomy improves glucose metabolism on gastric cancer patients: a nationwide population-based study. Surg Obes Relat Dis 2016;12(3):635-641. [doi: 10.1016/j.soard.2015.11.024] [Medline: 27012876]
21. Jin HY, Park TS, Lee KA, Baek YH. The influence of total or sub-total gastrectomy on glucose control in diabetic and non-diabetic patients. Acta Endocrinol (Buchar) 2016;12(4):423-430. [doi: 10.4183/aeb.2016.423] [Medline: 31149126]
22. Özdaş S, Olt S, Şirik M. The role of sleeve gastrectomy on preventing type 2 diabetes mellitus. Open Access Maced J Med Sci 2017 Jun 15;5(3):316-318. [doi: 10.3889/oamjms.2017.074] [Medline: 28698749]
23. Tran TB, Worhunsky DJ, Squires MH, et al. To Roux or not to Roux: a comparison between Roux-en-Y and Billroth II reconstruction following partial gastrectomy for gastric cancer. Gastric Cancer 2016 Jul;19(3):994-1001. [doi: 10.1007/s10120-015-0547-3] [Medline: 26400843]
24. Hoya Y, Mitsumori N, Yanaga K. The advantages and disadvantages of a Roux-en-Y reconstruction after a distal gastrectomy for gastric cancer. Surg Today 2009;39(8):647-651. [doi: 10.1007/s00595-009-3964-2] [Medline: 19639429]
25. Kim JH, Huh YJ, Park S, et al. Multicenter results of long-limb bypass reconstruction after gastrectomy in patients with gastric cancer and type II diabetes. Asian J Surg 2020 Jan;43(1):297-303. [doi: 10.1016/j.asjsur.2019.03.018] [Medline: 31060769]
26. Wei ZW, Li JL, Wu Y, et al. Impact of pre-existing type-2 diabetes on patient outcomes after radical resection for gastric cancer: a retrospective cohort study. Dig Dis Sci 2014 May;59(5):1017-1024. [doi: 10.1007/s10620-013-2965-6] [Medline: 24318804]
27. Yun L, Zhiwei J, Junsheng P, Xiaobin W, Cancan X, Jieshou L. Comparison of functional outcomes between functional jejunal interposition and conventional Roux-en-Y esophagojejunostomy after total gastrectomy for gastric cancer. Dig Surg 2020;37(3):240-248. [doi: 10.1159/000501677] [Medline: 31614348]
28. Kwon Y, Jung Kim H, Lo Menzo E, Park S, Szomstein S, Rosenthal RJ. A systematic review and meta-analysis of the effect of Billroth reconstruction on type 2 diabetes: a new perspective on old surgical methods. Surg Obes Relat Dis 2015;11(6):1386-1395. [doi: 10.1016/j.soard.2015.01.001] [Medline: 25892345]
29. Costa M, Trovão Lima A, Morais T, et al. Does reconstruction type after gastric resection matters for type 2 diabetes improvement? J Gastrointest Surg 2020 Jun;24(6):1269-1277. [doi: 10.1007/s11605-019-04255-4] [Medline: 31140062]
30. Maksimaityte V, Bausys A, Kryzauskas M, et al. Gastrectomy impact on the gut microbiome in patients with gastric cancer: a comprehensive review. World J Gastrointest Surg 2021 Jul 27;13(7):678-688. [doi: 10.4240/wjgs.v13.i7.678] [Medline: 34354801]
31. Tseng CH, Lin JT, Ho HJ, et al. Gastric microbiota and predicted gene functions are altered after subtotal gastrectomy in patients with gastric cancer. Sci Rep 2016 Feb 10;6:20701. [doi: 10.1038/srep20701] [Medline: 26860194]
32. Qi YF, Sun JN, Ren LF, et al. Intestinal microbiota is altered in patients with gastric cancer from Shanxi Province, China. Dig Dis Sci 2019 May;64(5):1193-1203. [doi: 10.1007/s10620-018-5411-y] [Medline: 30535886]

33. Lederer AK, Pisarski P, Kousoulas L, Fichtner-Feigl S, Hess C, Huber R. Postoperative changes of the microbiome: are surgical complications related to the gut flora? a systematic review. BMC Surg 2017 Dec 4;17(1):125. [doi: 10.1186/s12893-017-0325-8] [Medline: 29202875]
34. Guyton K, Alverdy JC. The gut microbiota and gastrointestinal surgery. Nat Rev Gastroenterol Hepatol 2017 Jan;14(1):43-54. [doi: 10.1038/nrgastro.2016.139] [Medline: 27729657]
35. Nomura E, Lee SW, Kawai M, et al. Functional outcomes by reconstruction technique following laparoscopic proximal gastrectomy for gastric cancer: double tract versus jejunal interposition. World J Surg Oncol 2014 Jan 27;12:20. [doi: 10.1186/1477-7819-12-20] [Medline: 24468278]
36. Nomura E, Kayano H, Lee SW, et al. Functional evaluations comparing the double-tract method and the jejunal interposition method following laparoscopic proximal gastrectomy for gastric cancer: an investigation including laparoscopic total gastrectomy. Surg Today 2019 Jan;49(1):38-48. [doi: 10.1007/s00595-018-1699-7] [Medline: 30159780]
37. Namikawa T, Munekage E, Munekage M, et al. Reconstruction with jejunal pouch after gastrectomy for gastric cancer. Am Surg 2016 Jun;82(6):510-517. [Medline: 27305882]
38. Namikawa T, Oki T, Kitagawa H, Okabayashi T, Kobayashi M, Hanazaki K. Impact of jejunal pouch interposition reconstruction after proximal gastrectomy for early gastric cancer on quality of life: short- and long-term consequences. Am J Surg 2012 Aug;204(2):203-209. [doi: 10.1016/j.amjsurg.2011.09.035] [Medline: 22813641]
39. bougtoir/gastric_dm. GitHub. URL: https://github.com/bougtoir/gastric_dm [accessed 2024-08-08]

Abbreviations

DM: diabetes mellitus
LJIP: laparoscopic jejunal interposition reconstruction
OT: other surgical techniques
RY: Roux-en-Y
SG: sleeve gastrectomy
TG: total gastrectomy

Edited by T Leung; submitted 15.01.24; peer-reviewed by V Fairhurst, J Olivier, O Oladoyin, MJC Machado, FQ Arogundade, M Noushad, R Jiwani; revised version received 19.06.24; accepted 20.06.24; published 14.08.24.

Please cite as:
Onishi T
Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study
JMIRx Med 2024;5:e56405
URL: https://xmed.jmir.org/2024/1/e56405
doi:10.2196/56405

Peer Review of "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study"

Femi Qudus Arogundade[1]; Syeda Azra[2], BSc; Myron Pulier[3], MD; Limegreen Ram; Daniela Saderi, PhD; Magdalena Tomaskova[4], MSc

1

2

3

4

Related Articles:

http://preprints.jmir.org/preprint/52970

https://med.jmirx.org/2024/1/e58859

https://med.jmirx.org/2024/1/e52970

(JMIRx Med 2024;5:e59430) doi:10.2196/59430

KEYWORDS

prosocial behavior; COVID-19; anxiety; COVID-19 prevention; preventive health behavior; adherence to prevention

This is the peer-review report for "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study."

Round 1 Review

This review is the result of a virtual collaborative live review discussion organized and hosted by PREreview and JMIR Publications. The discussion was joined by 16 people: 1 author, 2 facilitators, 2 members of the JMIR Publications team, and 11 live review participants. Aishah Ibrahim and Ananya Ananthakrishnan wished to be recognized for their participation in the live review discussion, even though they have not contributed to authoring the review below. We thank all participants who contributed to the discussion and made it possible for us to provide feedback on this preprint.

Summary

The study [1] investigates the complex interplay between anxiety (both state and trait), prosocial behaviors, and adherence to COVID-19 preventive measures among college students. While overall prosocial behaviors did not directly correlate with anxiety, a seemingly significant crossover effect emerged in relation to public prosocial behaviors, suggesting that individuals with lower self-oriented tendencies exhibited increased adherence behaviors under heightened state anxiety. The study used a quantitative research design with a sample of 54 undergraduate students, using online questionnaires to measure various psychological factors and preventive behaviors. Individuals with high anxiety showed increased adherence to preventive measures, contrary to the hypothesized moderating effect of prosocial behaviors. Overall, the reviewers appreciated the effort and recognized the challenges of conducting a research study in the context of an unprecedented social condition. However, the findings are challenged by weak effect sizes, multiple comparisons, and unclear appropriateness of using prosocial behaviors as a moderating variable. The study underscores the psychological impact of the COVID-19 pandemic on college students and suggests the need for further exploration into the nuanced relationships between anxiety, prosocial behaviors, and adherence to public health guidelines. Despite its strengths in data collection and questionnaire use, limitations such as a narrow participant pool and reliance on self-reporting warrant cautious interpretation of the results. The study encourages future research to delve deeper into these intricate connections, offering insights into potential interventions for promoting adherence to COVID-19 preventive measures and beyond.

Below we list major and minor concerns that were discussed by participants of the live review, and where possible, we provide suggestions on how to address those issues.

List of Major Concerns and Feedback

Small Sample Size and Mediation Analysis

One of the main concerns raised in the discussion was the small number of study participants. This is acknowledged as a limitation factor in the discussion, but what is less clear is if a mediation analysis is the right approach to analyze these data. One reviewer suggested the use of a multivariate analysis instead as it would take all variables into account without forcing

potentially artificially generated causal mediations between variables that don't show an obvious causal dependency. Another reviewer, however, felt that while this approach may work and it would be useful to explore, using a multivariate analysis may lead to overfitting in most covariates given the small sample size and given that 86% of subjects reported anxiety. Overall, the suggestion is for the authors to provide a rationale for selecting anxiety as a mediator variable over prosocial tendencies, or *vice versa*, and possibly explore other analyses and comment on the limitations of the approaches.

Uniform/Convenience Sampling

The reviewers acknowledged the study's challenging circumstances and the effort to capture unique data. However, they expressed concern about generalizability, noting that all participants were undergraduates from one college. The demographic homogeneity of this group may limit applicability to diverse populations, raising caution about extrapolating results across age groups, educational backgrounds, and cultural contexts. The convenience sampling method may have introduced bias, as easily accessible participants might not represent the broader target population.

University Policy and Mask Mandates

The study doesn't explicitly consider the potential impact of university campus policies during the pandemic, such as mask mandates and social distancing, on adherence behaviors. The findings may be influenced by the specific characteristics of the chosen university, including its restriction policies. For example, students on campus may wear face masks due to safety requirements when entering campus common spaces versus through their own personal decision process. Therefore, exercising caution in generalizing conclusions to broader contexts is suggested. Reviewers highlight the importance of future research with more diverse samples to enhance external validity, reinforcing the study's overall robustness. This provides a constructive pathway for refining the study's scope and applicability.

The study's cross-sectional design and reliance on self-reporting introduce potential limitations in establishing causal relationships and accurate data collection. (In general, no retrospective exploratory study can show causality, asserting a causal relationship amounts to the post hoc ergo propter hoc logical fallacy.) The one-time nature of the study also limits insights into the dynamic nature of psychological factors and preventive behaviors over time. Caution should be exercised in interpreting the results, as correlation does not imply causation. Furthermore, reviewers don't think that the results of this study can be used on their own to make any definitive public health policy recommendations.

Adherence Scores

The study mentions adherence to COVID-19 public health safety recommendations as an outcome variable. There is a need for more clarity on how the adherence scores were calculated, especially considering potential confounding factors such as university campus policies during the pandemic.

Missing Data

The study's approach to handling missing data in nonmandatory survey questions is not explicitly discussed. This may impact the results and should be clarified.

Reliability Metrics

The study does not provide information on test-retest reliability, accuracy against a gold standard, or error of measurement for the Prosocial Tendencies Measure (PTM) reliability. Reliability induction from other studies is mentioned, but the study population's specific reliability is not demonstrated. Without these critical reliability metrics, the study leaves a gap in the assessment of the psychometric properties of the PTM. Including such information would enhance the transparency and credibility of the study's findings, allowing readers to better evaluate the reliability and validity of the instrument used to assess prosocial behaviors. Future research may consider providing a comprehensive assessment of the psychometric properties of measurement instruments to strengthen the methodological rigor and overall quality of the study.

Statistical Model and Data Selection

Some reviewers expressed concern related to the lack of transparency about how variables were selected as moderators or mediators, how some others (eg, age) were chosen to be excluded, and how others were chosen to be reported on from the cited "larger study." Adding clarity around the rationale that led to making such choices would help the reader better contextualize the results.

Furthermore, a scoring guide for the CIS Survey would be helpful to add. There is a concern that a simple sum method may be biased because some questions may not be relevant to all subjects (eg, playdates only impact subjects that have childcare responsibilities).

Ethics

While the study mentions obtaining institutional review board approval and online passive consent, specific details regarding confidentiality, privacy safeguards, and participant understanding of risks are not thoroughly addressed.

Furthermore, it is not clear what the authors mean by "passive consent." Ordinarily, the term involves a parent or guardian giving consent on behalf of someone deemed not competent to give consent (see Range L, Embry T, MacLeod T. Active and passive consent: a comparison of actual research with children. *Ethical Hum Sci Serv*. 2001 Spring;3(1):23 - 31. PMID: 15278986). Were participants fully aware of what they were getting into, or does "passive" imply that consent was assumed by virtue of participants consenting to the terms of the "larger study"?

Data and Reproducibility

The study provides a moderate level of detail, but more specific information is needed for reproducibility. This includes additional demographic details, exact questionnaire wording, and more details on moderation analyses. The study would benefit from providing a more comprehensive set of demographic information about the participants such as age

distribution, gender distribution, and other relevant characteristics. A richer demographic profile would contribute to a more nuanced understanding of the study population and facilitate comparisons with other research. Reviewers suggested adding available details to Table 1.

While the study mentions that data are available upon reasonable request, reviewers suggest considering providing additional information on how interested researchers can request the data, perhaps from the corresponding author or another designated contact. This could enhance transparency and facilitate potential collaborations or further scrutiny of the results.

List of Minor Concerns and Feedback

Readability

Overall, the reviewers thought that the manuscript would benefit from a clearer explanation of key terms and recommended keeping the terminology consistent across the manuscript so as to help the reader better follow the narrative and interpret the findings. For example, there was some confusion among reviewers on the meaning of "public prosocial scale."

Approach and Results

It may be helpful to show more information about some of the background variables. One question is if the deviation of age from a normal distribution is significant and thus a possible contributor to the study's findings if age correlates with adherence or anxiety. Showing not only mean and SD but also median, quartiles, and range may provide a better feel for what the study population, or at least the participant sample, is like.

It may be useful to make explicit the assumptions underlying the modeling and parameters used for PROCESS, such as the degree of independence of the moderator.

Discussion

The authors may consider adding a section to the discussion to explore variables related to vaccine hesitancy and other factors (eg, sense of invincibility) as a suggestion for future research, expanding the scope beyond adherence to preventive measures.

Given the reliance on self-report measures, the reviewers suggest the authors discuss the potential impact of social desirability bias on participants' responses. Addressing this concern would add transparency to the limitations of the study.

Reviewers suggest authors discuss how the results support following up further on correlations among PTM scales and on the possible moderator effect of public prosocial tendencies, with recommendations for including a broader set (explicitly listed) of potentially explanatory independent variables.

It may also be helpful to add some explanation of why the psychometric characteristics of the survey instruments as established in other studies can be trusted to be the same as used in this study (online, unsupervised, etc). Some reviewers found it concerning that this study found statistically significant pairwise associations between PTM subscales, and this should be addressed, perhaps with speculation about why this happened.

Figures and Tables

Consider using a 2 × 2 table in Figure 1 to illustrate the detected moderator effect.

Title

Given the concern about generalizability, a reviewer suggested the authors consider changing the title to "Adherence Behaviors to Prevent COVID: The Role of Anxiety and Prosocial Behaviors Amongst University Students in the US."

Concluding Remarks

We thank the authors of the preprint for posting their work openly and for graciously agreeing to have their work reviewed via this process. We also thank all participants of the Live Review call for their time and for engaging in the lively discussion that generated this review.

Editorial Notice

This paper was peer-reviewed by the Plan P Hashtag Community partner #PeerRef.

Conflicts of Interest

DS helped condense the review notes and edit the final review. She is also the director of PREreview. All other authors declare no conflicts of interest.

Reference

1. Corbera S, Marin-Chollom AM. The role of anxiety and prosocial behaviors on adherence behaviors to prevent COVID-19 in university students in the United States: cross-sectional study. JMIRx Med 2024;5:e52970. [doi: 10.2196/52970]

Abbreviations

PTM: Prosocial Tendencies Measure

Edited by T Leung; submitted 11.04.24; this is a non–peer-reviewed article;accepted 11.04.24; published 27.05.24.

Please cite as:
Arogundade FQ, Azra S, Pulier M, Ram L, Saderi D, Tomaskova M
Peer Review of "The Role of Anxiety and Prosocial Behaviors on Adherence Behaviors to Prevent COVID-19 in University Students in the United States: Cross-Sectional Study"
JMIRx Med 2024;5:e59430
URL: https://xmed.jmir.org/2024/1/e59430
doi:10.2196/59430

Peer Review of "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study"

Vanessa Fairhurst; James Olivier[1]; Olajumoke Oladoyin[2]; Maria J C Machado; Femi Qudus Arogundade[3]; Mohammed Noushad; Rozmin Jiwani, PhD, RN

1

2

3

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.01.13.24301276v1

Companion article: https://med.jmirx.org/2024/1/e63859

Companion article: https://med.jmirx.org/2024/1/e56405

(JMIRx Med 2024;5:e63862) doi:10.2196/63862

KEYWORDS

diabetes mellitus; gastrectomy; gastric cancer surgery; glucose metabolism; postoperative diabetes onset; surgery outcomes

This is the peer-review report for "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study."

Round 1 Review

This review is the result of a virtual, collaborative live review discussion organized and hosted by PREreview and JMIR Publications. The discussion was joined by 18 people: 2 facilitators, 4 members of the JMIR Publication team, and 12 live review participants. Konstantinos Georgiou, Maria Florencia Grande Ratti, and Naser Kamyari wished to be recognized for their participation in the live review discussion, even though they have not contributed to authoring the review below. We thank all participants who contributed to the discussion and made it possible for us to provide feedback on this preprint.

Summary

The study [1] compares the results of Roux-en-Y (RY) reconstruction with other surgical techniques (OT) to determine the incidence of postoperative diabetes in patients with gastric cancer who had undergone total gastrectomy. The Tokyo Metropolitan Bokutoh Hospital cohort of 715 patients from 2005 to 2019 was examined. The study finds a statistically significant difference in the incidence of postoperative diabetes between the RY and OT groups, with RY associated with a greater incidence, through careful data preprocessing and statistical analysis. The study does admit many limitations, though, such as the absence of a control group that did not undergo a gastric bypass and the lack of assessment of the role that lifestyle factors and genetic predisposition play in the development of diabetes. The study also suggests more investigation into the possible effects of laparoscopic jejunal interposition reconstruction on gut flora and postoperative outcomes.

This retrospective, single-center study analyzed electronic medical records, which used hemoglobin A_{1c} (HbA_{1c}) levels as a surrogate for the determination of diabetes status in patients. The study aimed to examine the incidence of new-onset diabetes in patients with gastric cancer who had undergone gastrectomy. Interestingly, the author presents the data via Kaplan-Meier curves, which describe a statistically significant difference, revealing that patients who had an RY reconstruction were more likely to develop new-onset diabetes than patients where surgical reconstruction was achieved via other techniques.

While the findings are interesting, it is essential to enhance the clarity of the study by providing additional information on the sampling methods, the determination of sample size, and a breakdown of the number of events in each group to enable an accurate understanding of study procedures and outcomes. Moreover, an analysis of patients at risk of diabetes before surgery would reduce potential confounding factors. This could be achieved by including a Cox proportional hazard regression to potentially provide more information on the impact of reconstruction methods for the risk of developing diabetes, while also accounting for other covariates. An explanation and breakdown of other reconstructive techniques (in the OT group) would improve the utility and external validity of this study. Additionally, the participants could have had other comorbidities that could affect the outcome. Therefore, a note on the inclusion criteria and exclusion criteria is necessary.

Below we list major and minor concerns that were discussed by participants of the live review, and where possible, we provide suggestions on how to address those issues.

List of Major Concerns and Feedback

1. There was no rationale provided for the choice between RY and OT. Were any guidelines followed, or was this at the discretion of the attending physician?

2. Due to the complex nature of postoperative diabetes development, it is crucial to take any confounding variables into consideration and provide a full description of any adjustments made.

3. The author should consider including appropriate covariates in the study to assess if they have a confounding effect on the study's result. For instance, is the author able to stratify the patients in terms of their risk of developing diabetes or include relevant information such as family history or concurrent metabolic syndrome?

4. The author should explicitly state the study's inclusion and exclusion criteria. Please consider giving more details on the comorbidities of the included participants. This could be summarized, or tools such as the Charlson Comorbidity Index could be used.

5. Sufficient details are not provided to allow the reproduction of the study; thus, we suggest you follow the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting. For example, there is content in the *Methods* section that should go in *Results*, such as the number of participants included and their baseline characteristics in Table 1. In the same way, information is missing in the *Methods* section, such as clear definitions of outcomes, statistical analysis, or sample size calculation.

6. As the cumulative risk of bias for this type of study design is moderate to high, please identify all the variables used in the model. Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Clarify the inclusion and exclusion criteria, together with follow-up time frames and intervals. As the patients underwent surgery between 2005 and 2019, may we assume that the shortest follow-up after surgery was 3 or 4 years?

7. Also, describe any efforts to address potential sources of bias and explain how the study size was arrived at. Namely, the distribution of the age and sex of the participants is not clear, as there appears to be a bias toward male participants. Refer to the SAGER (Sex and Gender Equity in Research) guidelines for details on conducting a sex-based analysis and disaggregating data according to sex.

8. Please report the regression model used to assess the associations between the explanatory variables and survival or time to event. How did the author handle learning effects and the changing and evolving surgical or clinical protocols over the long time frame of this retrospective analysis? Discuss the generalizability of this modeling approach, as well as the direction and magnitude of any potential bias.

9. Please include explicit information regarding the competing outcome (ie, mortality events) and justify why no other clinical factors other than HbA_{1c} levels were considered.

10. How did the author confirm if the patients were free of diabetes at the time of surgery and before? It would be appropriate if the author provided the baseline (at the time of surgery or before) HbA_{1c} values of the study participants in Table 1.

11. The discussion focused on a procedure that was not mentioned elsewhere or used in this study. Please clarify if this procedure is part of your recommendation for the clinical management of these patients in the future. Additionally, mention if future planned studies will address any stratification of patients for risk of new-onset diabetes mellitus prior to the surgery, or any analysis of gut microbiota before and after surgery.

List of Minor Concerns and Feedback

12. Were any validation techniques used to verify the accuracy of the applied algorithms and analysis, such as code review, unit testing, or cross-validation?

13. It would be helpful to include a figure explaining the methodology, include more information about the proportion of different reconstructive techniques, and discuss results from other studies to attempt some comparisons for identifying what could have caused similarities or differences in this analysis. For instance, we do not know if the analysis of the groups was blinded.

14. The *Methods* section lacks proper referencing of previous studies to justify the choice of reconstruction methods (RY vs OT) and the criteria used for defining the onset of diabetes. Referencing previous studies that have investigated similar surgical techniques or criteria for diabetes onset would provide the necessary context and justification for the methods used in the study. Additionally, citing relevant literature would enhance the credibility of the study by demonstrating that the research methodology is grounded in established practices and informed by prior research findings.

15. Clear visualization of censored data points on a Kaplan-Meier survival curve is essential for accurately interpreting the survival probabilities and understanding the impact of censoring on the analysis. Optionally, you can include CIs for the stratified number of participants.

16. Due to the long time frame of the retrospective analysis and the possibilities of changes in protocol, the author should consider describing how learning effects were handled in the study.

17. There should have been more information about the American Society of Anesthesiologists score; this is a subjective score, so even if it was lifted from the electronic record, there ought to be a note pertaining to how many operators assigned the score and the degree of agreement between them.

18. It is unclear how the missing values were handled. Were they imputed based on a model? What was the definition of an outlier here: greater than 2.5 SDs? What data types are being

referred to here? And what inconsistencies needed to be corrected?

19. What happened to the study participants after 2008 in the OT group (Figure 1)? Why is there a straight line?

20. Please provide more detailed information on what the code does in this study and how it could be used elsewhere.

21. In the *Abstract*, the study setting has been indicated as "Electrical medical records." It should be "Electronic medical records."

Concluding Remarks

We thank the author of the preprint for posting their work openly for feedback. We also thank all participants of the live review call for their time and for engaging in the lively discussion that generated this review.

Conflicts of Interest

None declared.

Reference

1. Onishi T. Incidence of postoperative diabetes mellitus after Roux-en-Y reconstruction for gastric cancer: retrospective single-center cohort study. JMIRx Med 2024;5:e56405. [doi: 10.2196/56405]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
OT: other surgical techniques
RY: Roux-en-Y
SAGER: Sex and Gender Equity in Research
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by T Leung; submitted 01.07.24; this is a non–peer-reviewed article;accepted 01.07.24; published 14.08.24.

Please cite as:
Fairhurst V, Olivier J, Oladoyin O, Machado MJC, Arogundade FQ, Noushad M, Jiwani R
Peer Review of "Incidence of Postoperative Diabetes Mellitus After Roux-en-Y Reconstruction for Gastric Cancer: Retrospective Single-Center Cohort Study"
JMIRx Med 2024;5:e63862
URL: https://xmed.jmir.org/2024/1/e63862
doi:10.2196/63862

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org