JMIRx Med

Overlay journal for preprints with post-review manuscript marketplace
Volume 6 (2025) ISSN 2563-6316 Editor in Chief: Edward Meinert, MA (Oxon), MSc, MBA, MPA, PhD, CEng, FBCS, EUR ING

Contents

Peer-Review Reports

Authors’ Response To Peer Reviewss

Original Papers

Protocol

Peer Review of “Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis”

Anonymous

Related Articles:

Companion article: https://www.biorxiv.org/content/10.1101/2023.06.21.545938v2

Companion article: https://med.jmirx.org/2025/1/e69894

Companion article: https://med.jmirx.org/2025/1/e50458

(JMIRx Med 2025;6:e69895) doi:10.2196/69895

KEYWORDS

multistep fermentation; specific methane production; anaerobic digestion; kinetics study; biochar; first-order; modified Gompertz; mass balance; waste management; environment sustainability

This is a peer-review report for “Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis”

Round 1 Review

The present manuscript [1] deals with the study of the valorization of organic fractions of municipal solid waste through the production of volatile fatty acids and biogas. The article is interesting; in my opinion, it should be revised.

Comments

1. The presentation of the manuscript is very poor; the figures are not in the same format.
2. Some of the recent works should be discussed and cited in the Introduction section: [2-6].
3. The novelty of the work should be highlighted.
4. Full stops should be removed from all subheadings.
5. The Results and Discussion should be written in detail with proper subheadings.
6. There are some typo errors; they should be rectified.

Conflicts of Interest

None declared.

References

1. Borhany H. Converting organic municipal solid waste into volatile fatty acids and biogas: experimental pilot and batch studies with statistical analysis. JMIRx Med 2025;6:e50458. [doi: 10.2196/50458]
2. Jung S, Shetti NP, Reddy KR, et al. Synthesis of different biofuels from livestock waste materials and their potential as sustainable feedstocks – a review. Energy Conversion Manage 2021 May 15;236:114038. [doi: 10.1016/j.enconman.2021.114038]
3. Srivastava RK, Shetti NP, Reddy KR, Aminabhavi TM. Sustainable energy from waste organic matters via efficient microbial processes. Sci Total Environ 2020 Jun 20;722:137927. [doi: 10.1016/j.scitotenv.2020.137927] [Medline: 32208271]
4. Sampath P, Brijesh, Reddy KR, et al. Biohydrogen production from organic waste – a review. Chem Eng Technol 2020 Jul;43(7):1240-1248. [doi: 10.1002/ceat.201900400]
5. Velvizhi G, Goswami C, Shetti NP, Ahmad E, Kishore Pant K, Aminabhavi TM. Valorisation of lignocellulosic biomass to value-added products: paving the pathway towards low-carbon footprint. Fuel (Lond) 2022 Apr 1;313:122678. [doi: 10.1016/j.fuel.2021.122678]
6. Monga D, Shetti NP, Basu S, et al. Engineered biochar: a way forward to environmental remediation. Fuel (Lond) 2022 Mar 1;311:122510. [doi: 10.1016/j.fuel.2021.122510]

Edited by T Leung; submitted 10.12.24; this is a non–peer-reviewed article;accepted 10.12.24; published 04.02.25.

Please cite as:
Anonymous
Peer Review of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis"
JMIRx Med 2025;6:e69895
URL: https://xmed.jmir.org/2025/1/e69895
doi:10.2196/69895

Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69537

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e69870) doi:10.2196/69870

KEYWORDS

artificial intelligence; machine learning; algorithm; model; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the peer-review report for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study.”

Round 1 Review

General Comments

This paper [1] construct a checklist to support the development and implementation of artificial intelligence (AI) in clinical settings. I only have some minor comments.

Minor Comments

1. Comparison with existing checklists: Please add a comparison with some of the existing checklists.

2. Inconsistency in the number of studies: The authors initially stated that they included 20 studies, but later mentioned 23. Please clarify.

3. Appendix visibility: The appendix is not visible.

4. Abbreviation consistency: The abbreviation “IQR” appears multiple times. Please ensure it is clearly defined and used consistently.

Conflicts of Interest

None declared.

Reference

1. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]

Abbreviations

AI: artificial intelligence

Edited by CN Hang, E Meinert, T Leung; submitted 10.12.24; this is a non–peer-reviewed article;accepted 10.12.24; published 20.02.25.

Please cite as:
Anonymous
Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”
JMIRx Med 2025;6:e69870
URL: https://xmed.jmir.org/2025/1/e69870
doi:10.2196/69870

Peer Review for “Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.09.28.23295699v1

Companion article: https://med.jmirx.org/2025/1/e72092

Companion article: https://med.jmirx.org/2025/1/e53276

(JMIRx Med 2025;6:e72144) doi:10.2196/72144

KEYWORDS

point-of-care ultrasonography; training program; acute respiratory failure; acute circulatory failure; emergency department

This is the peer-review report for “Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study.”

Round 1 Review

General Comments

This paper [1] researches an essential component of point-of care ultrasonography. As this modality is rapidly evolving, evaluation of the impact on patient management and outcomes as well as cost-effectiveness is essential. Both subjects discussed in the paper result in a highly relevant manuscript. Even though the authors discuss relevant subjects, there are some problems with the manuscript.

Specific Comments

Major Comments

1. The title of the manuscript suggests that the authors researched the impact of ultrasound after implementation. However, as stated in the Methods section, ultrasound is already used by senior physicians. Thus, the impact of ultrasound after implementation is not researched but rather the impact of ultrasound used by residents. I suggest that the authors clarify that this is a feasibility and impact study on the implementation of point-of-care ultrasound (POCUS) used by residents in the emergency department (ED) in the title and Abstract.

2. The authors state that patients were not included consecutively due to logistics in phase 2. This results in a high risk of bias in the included patients. Please include in the CONSORT (Consolidated Standards of Reporting Trials) diagram the number of patients that were eligible and were excluded based on exclusion criteria or missed.

3. It is unclear how many residents were performing the ultrasound examinations included in the analysis: the Methods section state that there was only 1 resident at the ED in both phases, while in the Results section, it states that there were 12 residents trained. Please clarify.

4. The authors state that they chose a before-and-after implementation to evaluate the effect of POCUS to avoid contamination. However, before the implementation, POCUS was already used by senior physicians, which raises the question if POCUS was indeed not used in phase 1 of the trial.

5. Interestingly, in the Discussion section, the author discussed that the publication of Msolli et al [2] did not find an improvement of diagnostic accuracy. It would be interesting to discuss why this is the case.

6. In the Discussion and Conclusion, it is suggested that the use of POCUS might lead to a decrease in hospital mortality. Since this is an observational study in which, just as the authors state, a diagnostic tool rather than a therapeutic intervention is researched, this is too rash to state. Please remove this from the Conclusion and Abstract.

Minor Comments

Overall

7. The authors provide results with IQR; however, no ranges are given. Please describe results as mean (SD) when data are normally distributed or median (25th percentile – 75th percentile) when data are not normally distributed.

8. Formatting of the full manuscript needs some attention. For example, in the Abstract, not all sentences start with a capital letter. Also, it is common in the English language to write number in full up to 20.

9. Please follow the author guidelines of the journal for reporting values and the structure of the manuscript.

Title Page

10. The authors state that a clinical trial registration was done. However, it seems that they refer to a registration by a medical ethical review board. Please provide a clinical trial registration or if not applicable, remove it from the title page.

Introduction

11. In the first sentence, please state the full name of "emergency department" before using the abbreviation ED.

Methods

12. Figure 1 should be formatted. The most common formatting is according to the CONSORT flow diagram.

Results

13. Please do not discuss the results in the Results section.

Discussion

14. Please end the Discussion section with the strengths and limitations. The secondary findings should be above the Strengths and Limitations section.

Round 2 Review

I would like to compliment the authors of their extensive changes to the manuscript. I have some minor comments.

Minor Comments

1. I would suggest changing the sentence "However, there is still no strong evidence that the diagnostic accuracy of POCUS translates into a clinically relevant difference in patient outcomes" in the Introduction, because you also do not provide strong evidence (I do not know if we ever could provide strong evidence). I would suggest that you focus it more on the fact that the impact of using POCUS in the management of patients in the ED is still relatively unknown.

2. I would suggest to start the Discussion section with a short summary of the key findings.

Conflicts of Interest

None declared.

References

1. Bieler S, Tagan D, Grosgurin O, Fumeaux T. Impact of a point-of-care ultrasound training program on the management of patients with acute respiratory or circulatory failure by in-training emergency department residents (IMPULSE): before-and-after implementation study. JMIRx Med 2025;6:e53276. [doi: 10.2196/53276]
2. Msolli MA, Sekma A, Marzouk MB, et al. Bedside lung ultrasonography by emergency department residents as an aid for identifying heart failure in patients with acute dyspnea after a 2-h training course. Ultrasound J 2021 Feb 9;13(1):5. [doi: 10.1186/s13089-021-00207-9] [Medline: 33559777]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ED: emergency department
POCUS: point-of-care ultrasound

Edited by E Meinert, A Schwartz; submitted 04.02.25; this is a non–peer-reviewed article;accepted 04.02.25; published 03.03.25.

Please cite as:
Anonymous
Peer Review for "Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study"
JMIRx Med 2025;6:e72144
URL: https://xmed.jmir.org/2025/1/e72144
doi:10.2196/72144

Peer Review for “Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.07.30.24311256v1

Companion article: https://med.jmirx.org/2025/1/e68769

Companion article: https://med.jmirx.org/2025/1/e66213

(JMIRx Med 2025;6:e69705) doi:10.2196/69705

KEYWORDS

indocyanine green; ICG; sentinel lymph node; breast cancer; breast; fluorescence; axillary lymph node mapping; NIR; surgical planning; near-infrared

This is the peer-review report for “Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review.”

Round 1 Review

General Comments

This paper [1] summarized the application value and existing problems of indocyanine green (ICG) in sentinel lymph node (SLN) biopsy of early breast cancer, which has positive significance for improving the accuracy of clinical SLN detection. This study has certain clinical value.

Specific Comments

Major Comments

1. Due to the high hardware requirements for the clinical application of ICG, the number of relevant studies in the search is relatively small. It is hoped that the author can search the recent relevant literature to improve the credibility of this review.
2. It is hoped that the author will analyze and compare the advantages and disadvantages of ICG and traditional SLN biopsy methods, so as to guide clinicians to adopt appropriate methods for appropriate patients.

Conflicts of Interest

None declared.

Reference

1. Kurdi F, Kurdi Y, Reshetov IV. Applications of indocyanine green in breast cancer for sentinel lymph node mapping: protocol for a scoping review. JMIRx Med 2024;5:e66213. [doi: 10.2196/66213]

Abbreviations

ICG: indocyanine green
SLN: sentinel lymph node

Edited by S Tungjitviboonkun; submitted 05.12.24; this is a non–peer-reviewed article;accepted 05.12.24; published 06.01.25.

Please cite as:
Anonymous
Peer Review for "Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review"
JMIRx Med 2025;6:e69705
URL: https://xmed.jmir.org/2025/1/e69705
doi:10.2196/69705

Peer Review of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis"

Dina Elsalamony, MSc

Department of Biotechnology, Institute of Graduate Studies & Research, University of Alexandria, Alexandria, Egypt

Related Articles:

Companion article: https://www.biorxiv.org/content/10.1101/2023.06.21.545938v2

Companion article: https://med.jmirx.org/2025/1/e69894

Companion article: https://med.jmirx.org/2025/1/e50458

(JMIRx Med 2025;6:e69896) doi:10.2196/69896

KEYWORDS

multistep fermentation; specific methane production; anaerobic digestion; kinetics study; biochar; first-order; modified Gompertz; mass balance; waste management; environment sustainability

This is a peer-review report for "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis."

Round 1 Review

General Comments

Generally, the manuscript [1] should be strictly improved in English language writing and corrected for all grammatical errors throughout the whole manuscript. The author has to use a uniform style of the English language, either American or British English. Further English assistance is particularly required. Many missing articles and a lot of grammatical and punctuation errors must be corrected in the manuscript as in the corrected abstract.

Specific Comments

This paper shows an important aspect of multiple fermentation steps for the complete utilization of municipal solid waste and conversion to useful products, which is highly recommended for circular economic sustainability worldwide. However, it needs some major revision and arrangement to allow for a better presentation of this valuable work.

Major Comments

Title

1. "Valorization of Organic Fraction of Municipal Solid Waste Through Production of Volatile Fatty Acids (VFAs) and Biogas" is a long title that should be shortened to be more concise with no abbreviations—more indicative. Suggested title: "Valorization of Organic Municipal Solid Waste for Volatile Fatty Acids and Biogas Production."

Abstract Section

2. Generally speaking, it must be more concise and specific.

3. Please clearly mention the take-home message and the main findings of the research.

4. The abstract is too long and lacks the main methodology and main experimental techniques that were carried out in this work. The author may add some hints about the main methods used before mentioning the main results.

Manuscript

5. Keywords: Words must be modified to be more informative and representative of the research interest and differ from the word in the manuscript title. Maybe add "Multi Step of Fermentation Process" or "Waste Management and Environment Sustainability."

6. Arrangement of the experimental work in the manuscript may be needed in the Results and Discussion accordingly.

7. There is a lack of figures to describe the main parameter optimization steps well. Please reformulate to describe some data using figures with error bars.

8. The SD and table footnotes with the number of replicates should be noted underneath all of the given tables.

9. A mechanistic in-detail discussion is required, not just comparing your results with the previous work; justify better.

10. In research articles, do not include any table comparing literature results; the author can discuss the main findings in the text itself, as in Table 5.

11. The Conclusions section is missing in the manuscript to summarize and point out the novelty and the main findings from the research.

12. Generally speaking, in academic writing, (1) abstracts do not include abbreviations, (2) avoid articles in the title (the, a, an), and (3) avoid keywords that exist in the title.

13. As a rule of thumb, no dots in titles or subtitles as in the Experimental section, Anerobic Pilot Unities, etc.

14. Multiple references should be merged, not written separately, as in “29, 30” and “23, 27”; the author may use the merge reference option in reference software.

15. The author may add numbers for all titles and subtitles accordingly all over the manuscript.

Minor Comments

16. The author should avoid general and well-known information, and be selective in the recent references used. May add one small paragraph to the Biological Waste Management and Environment Sustainability section.

17. The author should clarify the main aim of the work clearly in the last paragraph of the Introduction.

18. Do not use our, we, or us in academic writing.

19. The author may mention novel applications of VFA and biogas. Mention different novel sources of biogas production.

20. The author should mention the gas chromatography type, gas injection rate, column dimensions, and the used carrier gas in the main document.

21. The author did not mention that flushing with nitrogen or carbon dioxide took place in anaerobic digestion while feeding reactors and how the anaerobic conditions were maintained; please mention it clearly or add the references used for the methodology.

22. Organize titles all over the manuscript.

23. Generally, the subtitles are too generic; modify them to be more indicative and precise.

24. “unless Saturday and Sunday” in line 208 is not important information; the suggested word “daily” is enough.

25. “Unite”: Please correct.

26. Remove the grid lines in the figures.

27. The author has to mention the range used for the chemical oxygen demand method, and the original reference should be cited appropriately.

28. “As can be seen”: This statement is repetitive more than once in the Discussion, in lines 301, 315, and 423.

29. Figure 3 caption: mesophilic fermentation: Please specify which stage because both of the sequential steps were called mesophilic fermentation in Figure 1.

30. What is the rationale for comparing 3 days to 4.5 days for all the used systems; the author may justify why 4.5 days is better to complete with this hydraulic retention time in the rest of the experiments or describe the one variable at a time optimization method that is used to determine the significant factors and the insignificant one; mention them clearly. Also, use in the Discussion the terms “significant” and “insignificant” according to the obtained *P* value.

31. The author has to mention tables and figures in the text in their appropriate place.

Round 2 Review

This paper is greatly enhanced compared to the previous copy, and the author followed the previous comments precisely.

I recommend its publication. Thanks for allowing me to review this interesting work.

General Note

The Word file is the correct revised one, but the PDF seems to be the old version.

Conflicts of Interest

None declared.

Reference

1. Borhany H. Converting organic municipal solid waste into volatile fatty acids and biogas: experimental pilot and batch studies with statistical analysis. JMIRx Med 2025;6:e50458. [doi: 10.2196/50458]

Abbreviations

VFA: volatile fatty acid

Edited by T Leung; submitted 10.12.24; this is a non–peer-reviewed article;accepted 10.12.24; published 04.02.25.

Please cite as:
Elsalamony D
Peer Review of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis"
JMIRx Med 2025;6:e69896
URL: https://xmed.jmir.org/2025/1/e69896
doi:10.2196/69896

Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”

Saima Zaki

Department of Physiotherapy, School of Allied Health Sciences, Sharda University, Plot No. 32-34, Knowledge Park III, Greater Noida, India

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69537

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e70058) doi:10.2196/70058

KEYWORDS

artificial intelligence; machine learning; algorithm; model; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the peer-review report for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study.”

Round 1 Review

General Comments

This paper [1] presents the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF), a checklist intended to support the development and implementation of AI systems in health care settings. The framework is built on a comprehensive literature review and a modified Delphi study involving health care professionals globally. The manuscript addresses a significant gap in the integration of AI by emphasizing the importance of sociotechnical considerations alongside technical aspects.

Specific Comments

Major Comments

1. Clarity and structure: The manuscript could benefit from clearer explanations, particularly in the methodology section. The description of the Delphi study and literature synthesis is dense and may be difficult for readers to follow. Consider breaking down complex sentences and using more straightforward language.

2. Methodological rigor: The manuscript lacks details on the selection process for Delphi panelists and the exact methods used for data analysis. Transparency in these areas would significantly strengthen the paper. Additionally, clarify how the Delphi method was modified and the rationale behind these modifications.

3. Literature review and contextualization: The discussion section could benefit from a more critical comparison between the CASoF and existing frameworks. While the manuscript mentions other frameworks, it does not fully explore their limitations or how the CASoF overcomes these challenges. Expanding this discussion would provide a stronger justification for the CASoF’s novelty and utility.

4. Checklist practicality: While the checklist is comprehensive, its length and complexity may hinder practical adoption. Consider providing a condensed version for quick reference and include examples of how the checklist can be applied in real-world scenarios.

5. Ethical considerations and bias mitigation: The manuscript discusses ethical considerations but lacks specific strategies for addressing these issues within the CASoF. Expanding this discussion would enhance the manuscript’s comprehensiveness.

Minor Comments

6. Typographical and grammatical errors: There are minor typographical and grammatical errors throughout the manuscript that should be corrected. For instance, phrases like “revised and edited” could be simplified to “revised” for conciseness.

7. Tables and figures formatting: The tables summarizing the Delphi study results are helpful but could be more effectively formatted. Using shading or color coding to distinguish between different stages or domains would improve clarity and ease of interpretation.

8. Recent references: Some references in the manuscript are relatively old, which is less ideal for a rapidly evolving field like AI. Where possible, the manuscript should incorporate more recent literature to support its claims and demonstrate the ongoing relevance of the topic.

Conflicts of Interest

None declared.

Reference

1. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]

Abbreviations

AI: artificial intelligence
CASoF: Clinical Artificial Intelligence Sociotechnical Framework

Edited by CN Hang, E Meinert, T Leung; submitted 13.12.24; this is a non–peer-reviewed article;accepted 13.12.24; published 20.02.25.

Please cite as:
Zaki S
Peer Review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study"
JMIRx Med 2025;6:e70058
URL: https://xmed.jmir.org/2025/1/e70058
doi:10.2196/70058

Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69537

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e69869) doi:10.2196/69869

KEYWORDS

artificial intelligence; machine learning; algorithm; models; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the peer-review report for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study.”

Round 1 Review

The paper [1] presents the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF), a structured approach to guide the planning, design, development, and implementation of AI systems in health care settings. The framework is designed to address the gap between technical performance and sociotechnical factors that are essential for successful AI deployment in clinical environments.

The authors conducted a literature synthesis and a modified Delphi study involving global health care professionals to develop and refine the CASoF checklist. The checklist emphasizes the importance of considering the value proposition, data integrity, human-AI interaction, technical architecture, organizational culture, and ongoing support and monitoring, to ensure that AI tools are not only technologically sound but also practically viable and socially adaptable within clinical settings.

The study found that the successful integration of AI in health care depends on a balanced focus on both technological advancements and the sociotechnical environment of clinical settings. The CASoF represents a step forward in bridging this divide, offering a holistic approach to AI deployment that is mindful of the complexities of health care systems. The checklist aims to facilitate the development of AI tools that are effective, user-friendly, and seamlessly integrated into clinical workflows, ultimately enhancing patient care and health care outcomes.

The authors acknowledge some limitations of the study, such as the need for continuous refinement of the CASoF through iterative feedback and broader engagement with more stakeholders. Future research should aim to include an even wider array of perspectives, particularly from underrepresented regions and specialties, to enhance the framework’s comprehensiveness and applicability.

Overall, the paper provides a valuable contribution to the field of AI in health care by offering a practical and comprehensive approach to the development and implementation of AI systems in clinical settings.

Conflicts of Interest

None declared.

Reference

1. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]

Abbreviations

AI: artificial intelligence
CASoF: Clinical Artificial Intelligence Sociotechnical Framework

Edited by CN Hang, E Meinert, T Leung; submitted 10.12.24; this is a non–peer-reviewed article;accepted 10.12.24; published 20.02.25.

Please cite as:
Anonymous
Peer Review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study"
JMIRx Med 2025;6:e69869
URL: https://xmed.jmir.org/2025/1/e69869
doi:10.2196/69869

Peer Review of "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development"

Colin Rogerson, MD, MPH

Division of Pediatric Critical Care, Regenstrief Center for Biomedical Informatics, Indiana University School of Medicine, 705 Riley Hospital Drive, Indianapolis, IN, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.22.24303209v1

Companion article: https://med.jmirx.org/2025/1/e71098

Companion article: https://med.jmirx.org/2025/1/e57719

(JMIRx Med 2025;6:e71100) doi:10.2196/71100

KEYWORDS

childhood pneumonia; community-acquired pneumonia; machine learning; clinical decision support system; prognostic care decision

This is the peer-review report for "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development."

Round 1 Review

General Comments

The authors [1] have examined the medical records for 437 patients with pneumonia and created a machine learning–based classifier to determine which patients required transfer to a tertiary care center. This subject is interesting, as the predictive power of these novel statistical techniques is high and could improve the clinical care of these patients. The authors have done thorough work describing the statistical methods used in the preprocessing of the data and model development. My primary concerns in the manuscript are the lack of clinical application description, the lack of description of the time frame of the included data elements, and the lack of description regarding the patient population and outcome of interest. The following are my point-by-point comments.

Specific Comments

Major Comments

Abstract

- The authors use the term "case management" in the Abstract and several times in the manuscript. In this context, the authors' meaning is the decision for the escalation of care or patient transfer. However, in US-based hospital systems, case management has a different meaning, which includes largely transition to rehabilitation or nursing facilities, acquisition of home oxygen therapy, etc. I would recommend altering this term for comprehension to something like "escalation of care" or "patient triage."
- The primary outcome of interest should be included in the Abstract.
- As detailed in the Methods section, it is crucial to describe the time frame for the included variables, to know when the algorithm could be used in clinical practice.

Introduction

- As the goal of the algorithm in the study is to predict which patients will need transfer to tertiary care for increasing respiratory support, more of the Introduction should focus on the management of in-hospital pediatric pneumonia, challenges, and reasons for the escalation of care.
- I would recommend altering the sentence that describes pneumonia as easily preventable and treatable. Several of the most complicated cases in the intensive care unit are admitted with pneumonia.

Methods

- While great care is taken to describe the approach to data preprocessing, feature selection, and model development, I would recommend following the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis) guidelines [2], which are validated reporting recommendations for predictive models.
- Please provide more details regarding the hospital systems involved in this study. Are they large, academic centers or small, rural centers?
- For study inclusion, I am not familiar with the Integrated Management of Childhood Illness guidelines. Are these structured diagnostic codes captured in the electronic health record? Is it a computational phenotype?

- Please specify what is meant by “neonatal age.”
- Many of the variables included in the model are colinear. For example, age and weight are highly dependent on one another, and including both in the model can be detrimental. The feature selection methods may be able to discern this, but maybe not. I would recommend using only age and *z* score in the model.
- The time frames are not stated for the variables. For example, does “hypoxia” mean hypoxia at any time during the hospitalization? On hospital admission? In the first 12 hours? This information is vital to determine the usability of the entire model. If the model uses variables available during the entire hospitalization, the predictive ability will be high, but the usability will be low. A model that can predict right when a patient is transferred to a tertiary care center that the patient will be transferred is useless. However, a model that can predict on admission, or in the first 6 - 12 hours, that a patient will require transfer is incredibly helpful. Without knowing the time frame for these variables, we cannot assess how the model could be applied in clinical practice.
- Please provide clarity regarding the study outcomes. The primary outcome is described as whether the patient was referred to a tertiary care center or not. The next sentence describes “poor prognosis” as pediatric intensive care unit admission or oxygen/ventilation support. How is this outcome used? Is this a secondary outcome? Is this describing the reason for transfer? Please clarify.
- As stated in the TRIPOD guidelines, you should present the amount of missingness in your data. It appears you used imputation methods for missing data. It is helpful to describe the amount of missing data that was imputed and the method for imputation.

Results

- There is a glaring lack of information regarding your study population. Please provide a table describing patient characteristics including demographics and the variables you used in the algorithm. Also, please provide a comparison between the patients who were transferred to a tertiary care center and those who were not.
- In imbalanced datasets, it can be more useful to measure model performance using the area under the precision-recall curve rather than the standard area under the receiver operator characteristic curve. I would recommend adding this metric.

Discussion

- The Discussion, overall, focuses much more on the technical details of the data curation and model development than it does on the clinical application of the model. Much of the technical details presented are also clearly explained in the Methods section and then repeated in the Discussion. I would recommend substantial revision to the Discussion section to remove redundant information that is already contained in the Methods section, as well as the addition of how this model could be applied in a clinical setting to improve the care of patients with pneumonia.
- The Discussion contains no information regarding the limitations of the study. Please describe in detail the prominent limitations of the study. These should include the use of retrospective data, including only two centers, imbalanced data, challenges with clinical implementation of the model, etc.
- The Discussion, and other areas of the manuscript, mention disease prevention several times. The goal of this study has nothing to do with the prevention of pneumonia, only the treatment of pneumonia and the prevention of associated morbidity and mortality. Please revise.

Conclusion

- As it stands, the Conclusion is fairly long and does not focus only on the primary findings of the study. I would recommend trimming it to 2 - 3 sentences that focus only on the primary findings of the study, such as the feasibility of developing this type of predictive model and the potential applications of the model to clinical practice.

Minor Comments

Methods

- The authors describe that ensemble methods “significantly enhance the accuracy of classifications.” Please provide a reference for this statement.

Results

- Please provide numbers for those who met your primary outcome of interest (transfer to a tertiary care center).
- Please provide a description of the time frame for patient transfer, for those who were transferred.

Discussion

- It would be interesting to hear more regarding the use of this model in resource-limited settings and the benefits it could provide.

Round 2 Review

General Comments

The authors have conducted a single-center, retrospective study evaluating the derivation and performance of a machine learning model to predict the need for transfer to a higher level of care for childhood pneumonia. The authors were provided with a substantial amount of feedback on the original submission, and although the authors’ response is detailed and comments on how all concerns were adequately addressed, the resulting manuscript is lacking in many if not most of the requested changes. The revised manuscript remains confusing to the reader and bereft of some essential elements of standard study reporting, including a basic description of the patient population and details regarding the timing of variable collection and use in the model. Due to this lack of response to the initial reviewer feedback, I am recommending rejection of this manuscript. The following are my point-by-point critiques, many of which are similar to those in my original review.

Specific Comments

Abstract

- First sentence: Please revise it to "Pneumonia is the leading cause of preventable mortality for children under five years of age."
- Background: The terms "case management" and "disease prevention" are still used in the Abstract. In my initial review, I recommended revising these terms to improve study clarity, and although the authors stated in their response that they replaced these terms, they remain in the Abstract. As it stands, it is not immediately clear to the reader that the purpose of the study was to provide a tool to assist bedside clinicians to determine which patients are likely to require transfer of care to a higher-level facility for pediatric pneumonia.
- Methods: As it stands, it is confusing to the readers what was actually done in the study. It should be very apparent that the authors used a specific list of variables (please provide each in the Abstract) to predict the need for transfer to a larger institution using a specific type of machine learning model (ensemble). In the current version, this is difficult to discern.
- Results: I would be completely clear regarding the outcome your model is predicting. After reading the paper, it is understood that "pneumonia prognosis" and "severity" actually mean required transfer to a higher level of care, but it is unclear in the Abstract. I would explicitly state "predicted transfer to a higher level of care with 77% - 88% accuracy."

Introduction

- Second paragraph, fifth sentence: I would recommend revising it to "However, this preventable health problem continues to be a substantial cause of mortality, especially in underdeveloped countries and regions, due to the lack of equipment and trained human resources." There is no way to quantify it as "the most important cause of mortality."
- The term "case management" continues to be used in the Introduction, which decreases clarity for the reader.
- As recommended previously, I would be very specific in the Introduction that you are trying to create a tool to help bedside clinicians (typically non–intensive care physicians) decide when to transfer a patient with pneumonia to a higher level of care to prevent morbidity and mortality. As it stands, this is unclear.

Methods

- In my initial review, I asked the authors to clarify what is meant by neonatal age. In their response, they said they had revised the Methods to state specifically 28 days or fewer. However, in the first paragraph of the Methods, it continues to state "neonatal age." Please revise.
- For clarity, I would recommend restating your primary outcome to simply "required tertiary care referral." Having the outcome as severe versus nonsevere, which is defined as requiring tertiary care referral or not, adds an extra step to the thought process and can be confusing.
- One of my largest concerns in the initial manuscript was the timing of the variables. This is crucial when determining how useful the model could be. If the elements in Table 1 are measured on admission, or in the first 6 - 12 hours of admission, the model could be very useful for patient care. If the elements were measured at any point during the hospitalization, it becomes much less useful. My worry is that the model was developed based on the elements' presence at any point, meaning if the child had fever, cough, respiratory distress, and hypoxia at hour 48, then at hour 49 the model was able to predict the patient would need transfer, and the patient was transferred at hour 50—this is not helpful to clinicians. On the other hand, if the model predicts at hour 12 that a patient needs transfer, and then at hour 50 they transfer, that is potentially very helpful to clinicians. Without these details, I cannot recommend the publication of the manuscript.
- It appears that the model was developed using the data from all 437 patients, and the results are presented following k-fold cross validation. It is standard practice to derive the model on a subset of the data (typically 70% - 80%) and then to test it on the remainder of the dataset to prevent overfitting and inflation of performance metrics. It does not appear that this was done. Despite having a small sample size, I believe this approach would lead to a more robust and generalizable model.

Results

- The first paragraph contains many "nuts and bolts" details of model development, and these would be better positioned in the Methods section.
- Both reviewers on the initial submission requested additional details describing the study population, and although the authors responded that they added these details, there are still none provided. It is essential to the understanding of the study results to know the characteristics of the patient population, and it should be a standard requirement for all clinical studies.
- The Shapley additive explanations value results presented in Figure 2 are valuable, but more details describing each measured factor are required. I recommend a table with each factor as rows and two columns comparing the population that did not require transfer to a tertiary care center to the population that did.
- An additional figure showing an area under the precision-recall curve for each model would also be interesting to the readers.

Discussion

- The Discussion spends a decent amount of space discussing the COVID-19 pandemic. While this does have some bearing on the management of childhood pneumonia, I believe the space would be better spent discussing the actual implementation of this type of algorithm. How would a primary care clinician actually use this model in practice? How would it improve upon current clinical practice? Would it be easy or difficult to incorporate into routine workflows? This would be more interesting to the readers.

- I recommend adding what the next steps of this line of research would be. How would you seek to improve the model's performance? More patient data? Additional variables?
- In the original submission, I recommended the authors provide a limitations section and also provided some examples. Although the authors response says they added this, there are still no limitations provided. Please provide this essential element to the Discussion.

Conclusion

- I recommend commenting on what the next steps of this line of research would be in more specific terms.

Round 3 Review

General Comments

The authors have conducted a single-center, retrospective study evaluating the derivation and performance of a machine learning model to predict the need for transfer to a higher level of care for childhood pneumonia. The authors were provided with a substantial amount of feedback on the original submission and have been responsive to feedback, which has resulted in a much improved manuscript. There remain several typographical and grammatical errors, which I would advise an English-grammar expert to review prior to publication, but from a scientific standpoint, I believe the manuscript is appropriate for publication.

Specific Comments

Major Comments

1. Details regarding the patient population have been provided in detail.

2. The study objectives have been clarified for readers.

3. The study methods are now much more reproducible.

Conflicts of Interest

None declared.

References

1. Serin O, Akbasli IT, Cetin SB, et al. Predicting escalation of care for childhood pneumonia using machine learning: retrospective analysis and model development. JMIRx Med 2025;e57719:6. [doi: 10.2196/57719]
2. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 7;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]

Abbreviations

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis

Edited by E Meinert, S Amal, T Leung; submitted 09.01.25; this is a non–peer-reviewed article;accepted 09.01.25; published 04.03.25.

Please cite as:
Rogerson C
Peer Review of "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development"
JMIRx Med 2025;6:e71100
URL: https://xmed.jmir.org/2025/1/e71100
doi:10.2196/71100

Peer Review of “Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study”

Ali Ahmed, MPhil, PhD, PharmD

Division of Infectious Diseases and Global Public Health, School of Medicine, University of California, San Diego, La Jolla, CA, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.03.24302286v1

Companion article: https://med.jmirx.org/2025/1/e71528

Companion article: https://med.jmirx.org/2025/1/e57597

(JMIRx Med 2025;6:e71529) doi:10.2196/71529

KEYWORDS

periodic health examination; PHE; preventive health services; routine health checkups; Jordan; cross-sectional study

This is a peer-review report for “Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study.”

Round 1 Review

Specific Comments

Major Comments

1. In this manuscript [1], write in detail about the data collection procedure.
2. Why was a convenience sampling technique employed?
3. “All collected data are treated with strict confidentiality.” Some language corrections are required.

Minor Comments

There are a lot of formatting issues; many things seem copied and pasted.

Conflicts of Interest

None declared.

Reference

1. Tayoun AA. Determinants of periodic health examination uptake: insights from a Jordanian cross-sectional study. JMIRx Med 2025;6:e57597. [doi: 10.2196/57597]

Edited by T Leung; submitted 20.01.25; this is a non–peer-reviewed article;accepted 20.01.25; published 05.02.25.

Please cite as:
Ahmed A
Peer Review of “Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study”
JMIRx Med 2025;6:e71529
URL: https://xmed.jmir.org/2025/1/e71529
doi:10.2196/71529

Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”

Keith Thompson, MD

Department of Family Medicine, Western University, 1151 Richmond St, London, Canada

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69537

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e69593) doi:10.2196/69593

KEYWORDS

artificial intelligence; machine learning; algorithm; model; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the peer-review report for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study.”

Round 1 Review

General Comments

This paper [1]...is a very cohesive approach to establishing a framework for the implementation of artificial intelligence (AI).

Specific Comments

Major Comments

1. Ideally there should be information on the demographics of the expert panel.

2. Please add comments regarding equitable access for these technologies.

Conflicts of Interest

None declared.

Reference

1. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]

Abbreviations

AI: artificial intelligence

Edited by CN Hang, E Meinert, T Leung; submitted 03.12.24; this is a non–peer-reviewed article;accepted 03.12.24; published 20.02.25.

Please cite as:
Thompson K
Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”
JMIRx Med 2025;6:e69593
URL: https://xmed.jmir.org/2025/1/e69593
doi:10.2196/69593

Peer Review of "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study"

Sanjeev Kumar Thalari, MBA, MSc, PhD

Department of Management Studies & Research Center, CMR Institute of Technology, Bangalore, India

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.04.12.23288461v1

Companion article: https://med.jmirx.org/2025/1/e70059

Companion article: https://med.jmirx.org/2025/1/e48346

(JMIRx Med 2025;6:e70808) doi:10.2196/70808

KEYWORDS

rural alimentation; community health workers; motivation; retention; rural health; rural nutrition; workforce

This is the peer-review report for "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study."

Round 1 Review

General Comments

This paper [1] has given the impression that the researcher has done thorough homework before starting the research and it is evident in the paper. Case methodology and thematic analysis are a few of the approaches that depict the quality of the paper. Overall, as a reviewer, it is my opinion that the research paper is of quality.

Specific Comments

1. A few more factors like government initiatives should be included in studying the impact on the motivation and retention of community health workers.

Major Comments

1. I feel that the analysis also can include education as a parameter.

2. The thematic analysis is one of the strengths of this research and is appreciated.

Minor Comments

1. Common wording should be used in every section of the paper, like qualitative case research methodology and qualitative case research.

Conflicts of Interest

None declared.

Reference

1. Kerketta A, A N R. The impact of rural alimentation on the motivation and retention of Indigenous community health workers in India: qualitative study. JMIRx Med 2025;6:e48346. [doi: 10.2196/48346]

Edited by A Schwartz; submitted 02.01.25; this is a non–peer-reviewed article;accepted 02.01.25; published 23.01.25.

Please cite as:
Kumar Thalari S
Peer Review of "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study"
JMIRx Med 2025;6:e70808
URL: https://xmed.jmir.org/2025/1/e70808
doi:10.2196/70808

Peer Review of “Mothers’ Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study”

Bilkisu Nwankwo, MSc

Department of Community Medicine, College of Medicine, Kaduna State University, Kaduna, Nigeria

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.04.05.24305403v1

Companion article: https://med.jmirx.org/2025/1/e70145

Companion article: https://med.jmirx.org/2025/1/e59379

(JMIRx Med 2025;6:e70142) doi:10.2196/70142

KEYWORDS

mother's knowledge and practices; oral hygiene; child oral health; bangladesh

This is the peer-review report for “Mothers’ Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study.”

Round 1 Review

Specific Comments

There were a lot of grammatical issues and typographical errors. The manuscript [1] needs to be edited for grammar and syntax. It is also obvious that the manuscript was not proofread adequately.

Major Comments

Abstract

- A word is missing in the first sentence. Authors should proofread the manuscript.
- Keywords: Dhaka is a more appropriate keyword than Bangladesh.
- Under the Results in the abstract, respondents should be referred to as such and not as samples.

Introduction

- The global prevalence of oral diseases was stated, but authors did not capture the prevalence in the study area/country and so have not shown that oral disease is a problem. Even the global prevalence that was stated was only that of dental caries among the seven conditions that make up oral diseases as stated by the authors.
- The objective stated here (last sentence) comes off like the authors are assessing the knowledge and practices of oral hygiene with regard to themselves and not their children as stated in the topic.

Methods

- Was it permission that was given by the institutional review board or an ethical clearance?
- This section is quite disorganized. There is a logical flow expected in this section.
- Why was a nonprobability sampling technique (convenient sampling) used for this study? The sampling technique was not explained at all. This will make replicating this study difficult.
- I have an issue with the scoring system and the grading. Is there a reference for it? I particularly have an issue with “moderately average.” It is not a standard term.
- The exclusion criteria are not the opposite of the inclusion criteria as stated by the authors. Exclusion criteria are those already included in the study but that are ineligible for one reason or the other.

Results

- In the text above Table 1, authors wrote that most respondents (39.3%) had a monthly family income of “21,000 - 40,000 taka per month.” This figure (39.3%) is just over one-third of the respondents and not a majority.
- Table 1: What is the meaning of graduation and above? Is it graduated secondary school or graduated college?
- “Respectively” should be added at the end of the following sentence. “Out of 400 mothers, more than 90% knew the importance of brushing teeth while 82.3% and 80.8% of them knew the recommended frequency and appropriate time for brushing teeth.”

Discussion

- The second sentence: the study is not evaluating parent’s knowledge and practices but that of mothers.
- Grammatical errors and missing words

Reference List

- Some of the references were not cited correctly. Authors should adhere to the Vancouver referencing style.

Conflicts of Interest

None declared.

Reference

1. Tamannur T, Das SK, Nesa A, et al. Mothers' knowledge of and practices toward oral hygiene of children aged 5-9 years in Bangladesh: cross-sectional study. JMIRx Med 2025;6:e59379. [doi: 10.2196/59379]

Edited by T Leung; submitted 16.12.24; this is a non–peer-reviewed article;accepted 06.01.25; published 03.02.25.

Please cite as:
Nwankwo B
Peer Review of "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study"
JMIRx Med 2025;6:e70142
URL: https://xmed.jmir.org/2025/1/e70142
doi:10.2196/70142

Peer Review of "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study"

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.03.24302286v1

Companion article: https://med.jmirx.org/2025/1/e71528

Companion article: https://med.jmirx.org/2025/1/e57597

(JMIRx Med 2025;6:e71531) doi:10.2196/71531

KEYWORDS

periodic health examination; PHE; preventive health services; routine health checkups; Jordan; cross-sectional study

This is a peer-review report for "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study."

Round 1 Review

The following items were noted in this paper [1].

- Periodic health examination (PHE) uptake: Only 27.1% of participants underwent a PHE in the last 2 years.
- Predictors: Significant predictors include recent visits to a primary health care facility, monthly income, and knowledge about PHEs and preventive health measures.
- Nonsignificant factors: Gender, marital status, smoking status, and BMI did not show a significant association with PHE uptake.

Strengths

1. Comprehensive analysis: The study employs a robust methodology, combining descriptive, inferential, and multivariate statistical techniques to provide a thorough understanding of PHE uptake.
2. Significant predictors identified: Key factors influencing PHE uptake were identified, offering valuable insights for health care providers and policy makers.
3. First of its kind in Jordan: This study fills a gap in existing knowledge by being the first to investigate PHE uptake in Jordan.

Negative Points and Areas for Improvement

Cross-Sectional Design

- Limitation: The study's design limits the ability to establish causality.
- Improvement: Future research could benefit from a longitudinal approach to better establish causal relationships between the identified predictors and PHE uptake.

Convenience Sampling

- Limitation: This method may introduce selection bias, and the online survey format may lead to measurement bias.
- Improvement: Employing a more randomized and stratified sampling method could enhance the representativeness and validity of the findings.

Limited Generalizability

- Limitation: Results may not be generalizable to populations outside of Jordan or those not included in the sample.
- Improvement: Expanding the study to include diverse populations and different geographic regions would provide a more comprehensive understanding of PHE uptake.

Survey Instrument

- Limitation: The questionnaire's comprehensiveness and relevance to the Jordanian context might not have been fully ensured.
- Improvement: Pretesting the survey with a larger and more varied group, followed by adjustments based on feedback, could improve its applicability and accuracy.

Behavioral Factors

- Limitation: The study did not find a relationship between behavioral factors and PHE uptake, which contradicts findings in other contexts.
- Improvement: A more detailed investigation into cultural and societal influences on health behaviors in Jordan is needed to clarify these results.

English Language and Clarity

- Limitation: The manuscript contains some grammatical errors and awkward phrasings, which can detract from its readability.
- Improvement: A thorough review and editing for language and clarity by a native English speaker or professional editor would enhance the manuscript's quality.

Conflicts of Interest

None declared.

Reference

1. Tayoun AA. Determinants of periodic health examination uptake: insights from a Jordanian cross-sectional study. JMIRx Med 2025;6:e57597. [doi: 10.2196/57597]

Abbreviations

PHE: periodic health examination

Edited by T Leung; submitted 20.01.25; this is a non–peer-reviewed article;accepted 20.01.25; published 05.02.25.

Please cite as:
Anonymous
Peer Review of "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study"
JMIRx Med 2025;6:e71531
URL: https://xmed.jmir.org/2025/1/e71531
doi:10.2196/71531

Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”

Sai Saripalli, MSc

Louisiana State University, Baton Rouge, LA, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69537

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e69594) doi:10.2196/69594

KEYWORDS

artificial intelligence; machine learning; ML; algorithm; model; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the peer-review report for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study.”

Round 1 Review

General Comments

Using artificial intelligence (AI) to add social and domain-specific steps to clinical trials is innovative [1]. My only comment is whether the number of stages or the checklist changes if the shortlisted panelists change.

Specific Comments

Major Comments

1. I am unsure if having 38 (expert) panelists is good enough to have a robust framework.

Conflicts of Interest

None declared.

Reference

1. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]

Abbreviations

AI: artificial intelligence

Edited by CN Hang, E Meinert, T Leung; submitted 03.12.24; this is a non–peer-reviewed article;accepted 03.12.24; published 20.02.25.

Please cite as:
Saripalli S
Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”
JMIRx Med 2025;6:e69594
URL: https://xmed.jmir.org/2025/1/e69594
doi:10.2196/69594

Peer Review of “Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.22.24303209v1

Companion article: https://med.jmirx.org/2025/1/e71098

Companion article: https://med.jmirx.org/2025/1/e57719

(JMIRx Med 2025;6:e71369) doi:10.2196/71369

KEYWORDS

childhood pneumonia; community-acquired pneumonia; machine learning; clinical decision support system; prognostic care decision

This is the peer-review report for “Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development.”

Round 1 Review

General Comments

This paper [1] developed a machine learning approach that could predict community-acquired pneumonia prognosis, which is scaled into two levels, severe or nonsevere, and identify important clinical indices, such as hypoxia, respiratory distress, age, *z* score of weight for age, and antibiotic usage before admission. The machine learning–based clinical decision support system tool for childhood pneumonia could provide prognostic support for case management.

Specific Comments

Major Comments

1. To enhance the manuscript’s grounding in current research and to provide a comprehensive context for the study, the authors are recommended to incorporate an evaluation of related literature in the Introduction and Discussion sections. This could include, but not be limited to, the following studies:

- Liu YC, Cheng HY, Chang TH, et al. Evaluation of the need for intensive care in children with pneumonia: machine learning approach. JMIR Med Inform. Jan 27, 2022;10(1):e28934. [doi: 10.2196/28934] [Medline: 35084358]
- Smith JC, Spann A, McCoy AB, et al. Natural language processing and machine learning to enable clinical decision support for treatment of pediatric pneumonia. AMIA Annu Symp Proc. Jan 25, 2020;2020:1130-1139. [Medline: 33936489]
- Kanwal K, Khalid SG, Asif M, Zafar F, Qurashi AG. Diagnosis of community-acquired pneumonia in children using photoplethysmography and machine learning-based classifier. Biomed Signal Process Control. Jan 2024;87:105367. [doi: 10.1016/j.bspc.2023.105367]
- Chang TH, Liu YC, Lin SR, et al. Clinical characteristics of hospitalized children with community-acquired pneumonia and respiratory infections: Using machine learning approaches to support pathogen prediction at admission. J Microbiol Immunol Infect. Aug 2023;56(4):772-781. [doi: 10.1016/j.jmii.2023.04.011] [Medline: 37246060]

The readers could have a more comprehensive understanding if the authors could include a concise evaluation of the prior literature in the current manuscript.

2. Considering the high stakes involved in pediatric care, particularly in intensive settings, it is critical to exam the false negative cases from the confusion matrices. Analyzing these cases for any common feature characteristics could provide insights into potential improvements in the predictive algorithm. This analysis should be clearly presented and discussed in the manuscript, emphasizing its importance in clinical decision-making.

3. The manuscript would benefit from a more detailed description of the cohort used in the study. Information on age, gender, and other clinical indices across the two groups (severe and nonsevere) would enable a better understanding of the study population. Additionally, providing the number of cases in each group would clarify the scope and scale of the study findings.

4. A detailed description of the data collection process is crucial for assessing the study’s applicability in real-world clinical settings. The manuscript should explicitly state the following:

- How and when clinical data, including features such as hypoxia and respiratory distress, were collected (eg, at the time of admission? or within 24 hours of admission?);

- The time frame considered for “antibiotic usage before admission” as relevant to the prediction model: This information is essential for replicability and for future applications of the findings in clinical workflows.

Round 2 Review

I thank the authors for revising the manuscript.

Conflicts of Interest

None declared

Reference

1. Serin O, Akbasli IT, Cetin SB, et al. Predicting escalation of care for childhood pneumonia using machine learning: retrospective analysis and model development. JMIRx Med 2025;6:e57719. [doi: 10.2196/57719]

Edited by E Meinert; submitted 16.01.25; this is a non–peer-reviewed article;accepted 16.01.25; published 04.03.25.

Please cite as:
Anonymous
Peer Review of “Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development”
JMIRx Med 2025;6:e71369
URL: https://xmed.jmir.org/2025/1/e71369
doi:10.2196/71369

Peer Review of “Mothers’ Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study”

Md Hafizul Islam

Institute of Nutrition and Food Science, University of Dhaka, Dhaka, Bangladesh

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.04.05.24305403v1

Companion article: https://med.jmirx.org/2025/1/e70145

Companion article: https://med.jmirx.org/2025/1/e59379

(JMIRx Med 2025;6:e70144) doi:10.2196/70144

KEYWORDS

mothers’ knowledge and practices; oral hygiene; child oral health; Bangladesh

This is the peer-review report for “Mothers’ Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study.”

Round 1 Review

This is an interesting piece of research [1], which highlights mothers’ knowledge and practices regarding their children’s oral health in Dhaka City. However, several issues made the study scientifically questionable. The major issues are as follows. The study included mothers from two hospitals in Dhaka City, but the title of the study does not mention this. The sample selection from the mothers visiting the hospitals might not represent general mothers from the whole of Dhaka. Thus, this study might not be generalizable to all mothers in Dhaka City.

Introduction

Revise the last paragraph of the Introduction to highlight the study gap in Bangladesh and clearly state the objective of the study. Use the formal word “mother” and avoid the word “moms.”

Methods

Study Setting and Participants

Give clear reasoning as to why you selected study participants from the hospitals. The last line is confusing. It is not clear whether the participants filled out the questionnaire on their own or they were interviewed by the enumerators.

Sampling Technique

Please mention the nonresponse bias for the convenient sampling. Give a short description of the pretesting mentioning the number of samples, period, and location for it.

Measurement of Knowledge and Practice Score

Give the 15 knowledge-related questions and 13 practice-related questions in the supplementary file. Mention if these questions are your own or if you used any valid tools or questions adopted from the relevant previous studies. Give adequate information regarding the scoring system of the variables, mentioning the highest possible aggregated score and examples of two questions (one for knowledge and one for practice).

Statistical Analyses

The authors mentioned that they used the Mann-Whitney *U* test and the Kruskal-Wallis test. However, they did not mention the underlying assumptions of the tests. Moreover, the Results section also shows the χ^2 test but is not mentioned in the Methods section. Furthermore, the last line of the Results of the abstract shows the Pearson correlation coefficient, but nothing is mentioned in the Methods or Results section of the entire manuscript.

Results

Table 1

It is confusing as the text description of Table 1 and the title of Table 1 are different. It is recommended to use two separate tables: one for socioeconomic variables and another for the frequency distribution of the knowledge level among socioeconomic variables. Mention the knowledge- and practice-related raw scores first and then the cross-tab results. There is a major mistake in the results of Tables 1 and 2. The frequency distribution for educational status, occupation, family type, number of family members, and monthly income in Tables 1 and 2 are the same. However, the *P* values are different. How is this possible? Please check the results.

Discussion

It is confusing whether the practice was for the children or how a mother takes care of their children's dental health. Mention the implications of your findings rather than just comparing the findings with previous studies. State the limitation of the study, especially the bias regarding convenient sampling. Provide a section on the public health significance of the study findings in Bangladesh.

Conclusion

The Conclusion section of the study is poorly written and not focused on the findings of the study. Revise the Conclusion section to highlight your study findings.

Round 2 Review

The authors impressively amended the initial version of the manuscript based on the reviewers' comments. However, several issues remain unaddressed.

1. The authors should include the city in the title of the study. You can revise the title to "Knowledge and practices towards oral hygiene of children aged 5 - 9 years old: a cross-sectional study among mothers visited tertiary level hospitals in Dhaka, Bangladesh."
2. Use the full form when it appears first and then use the abbreviation afterward. For example, "KP" in the abstract.
3. Please mention this statistical test in the Methods section of the abstract. You did not mention the χ^2 test and Pearson correlation.
4. It is recommended to make the recommendation simple and easy to understand for the readers. Avoid duplication of the same term.
5. In the sample size calculation, you used *P*=.58 and *P*=.57. Please clarify why you used those prevalences. Cite the relevant study here.
6. Before the heading for the sociodemographic variables in the Methods section, you mention outcome measures. However, the sociodemographic variables are not your outcome variables according to your objectives. You can remove the term outcome measures from here.
7. You mentioned that you used 13 questions for the assessment of practices. Thus, according to your scoring approach, there should be a score of 1-13, but here, it is 1-11.
8. Please mention the name of the software and version you used for the statistical analysis.
9. Revise the sentence before Table 1. You can make it two sentences. One for family income and another for occupation.
10. There is no chi-square–related data in Table 1. Please remove the footnotes from Table 1.
11. In Figure 1, it is recommended to keep the values to one decimal point for 1a and 1b.
12. Please revise the sentence before Table 3 to give a clear meaning.
13. You can remove the percentage symbol from the value and give it in the vertical axis title.
14. Please give the correlation results in the main manuscript or as a supplementary table.
15. The authors overlooked the association of knowledge and practice with income and family size. Please give more details on those two points in the Discussion section.

Conflicts of Interest

None declared.

Reference

1. Tamannur T, Das SK, Nesa A, et al. Mothers' knowledge of and practices toward oral hygiene of children aged 5-9 years in Bangladesh: cross-sectional study. JMIRx Med 2025;6:e59379. [doi: 10.2196/59379]

Edited by T Leung; submitted 16.12.24; this is a non–peer-reviewed article;accepted 16.12.24; published 03.02.25.

Please cite as:
Islam MH
Peer Review of "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study"
JMIRx Med 2025;6:e70144
URL: https://xmed.jmir.org/2025/1/e70144
doi:10.2196/70144

Peer Review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69537

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e69595) doi:10.2196/69595

KEYWORDS

artificial intelligence; machine learning; algorithm; model; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the peer-review report for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study.”

Round 1 Review

This paper [1] introduces the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF), a checklist developed through a literature synthesis and refined by a modified Delphi study. It aims to guide the development and implementation of AI in clinical settings, focusing on the integration of both technological performance and sociotechnical factors. The framework addresses gaps in existing frameworks by emphasizing not only technical specifications but also the broader sociotechnical dynamics essential for successful AI deployment in health care.

New approaches to reporting AI in clinical settings are crucial as AI becomes more integrated into clinical practice. However, the paper needs to address the “black box” dilemma more thoroughly. This refers to the opaque nature of AI algorithms, where the decision-making process is not easily interpretable by clinicians, leading to trust and transparency issues. Additionally, while the CASoF checklist is a valuable tool, it would benefit from a more detailed comparison to established frameworks like TRIPOD (Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis), which has been widely used in developing and validating clinical prediction models. Discussing how the CASoF complements or improves upon TRIPOD would strengthen the paper’s contributions.

I suggest adding a paragraph discussing the potential roles of AI when integrated into hospital electronic health record (EHR) systems. AI could be used for the development of advanced diagnostic and prognostic tools by analyzing real-time patient data. Integration with EHRs could enhance decision-making, providing predictive analytics at the point of care and improving patient outcomes. This would help explore the broader clinical impact of AI beyond just technical integration, addressing its potential for continuous learning and optimization in health care settings.

Conflicts of Interest

None declared.

Reference

1. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]

Abbreviations

AI: artificial intelligence
CASoF: Clinical Artificial Intelligence Sociotechnical Framework
EHR: electronic health record
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis

Edited by CN Hang, E Meinert, T Leung; submitted 03.12.24; this is a non–peer-reviewed article;accepted 03.12.24; published 20.02.25.

Please cite as:
Anonymous
Peer Review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study"
JMIRx Med 2025;6:e69595
URL: https://xmed.jmir.org/2025/1/e69595
doi:10.2196/69595

Peer Review of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis"

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.03.23292185v1

Companion article: https://med.jmirx.org/2025/1/e69307

Companion article: https://med.jmirx.org/2025/1/e50712

(JMIRx Med 2025;6:e70039) doi:10.2196/70039

KEYWORDS

breast; cancer; oncology; ovarian; virus; viral; Epstein-Barr; herpes; bioinformatics; chromosome; gene; genetic; genetics; chromosomal; DNA; genomic; BRCA; metastasis; biology

This is the peer-review report for "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis."

Round 1 Review

Review Report With Major Revisions for the Paper

Title: "Herpesvirus infections eliminate safeguards against breast cancer and its metastasis: comparable to hereditary breast cancers"

Summary

The paper [1] hypothesizes that Epstein-Barr virus (EBV) infections promote breast cancer by disabling cancer safeguards. It is a bioinformatics analysis of public information from about 2100 breast cancers. The study finds that breast and ovarian cancer breakpoints cluster around EBV-associated cancer breakpoints, suggesting a significant role of EBV in promoting these cancers. The paper also identifies similarities in the molecular and cellular disruptions caused by EBV with those found in hereditary breast cancers.

Major Revisions Needed

Clarification of Hypotheses and Objectives

The hypothesis, while intriguing, needs clearer articulation. Specifically, the connection between EBV and breast cancer needs more explicit theoretical underpinning. Clarify the objectives and expected outcomes of the study at the outset.

Methodological Rigor and Data Sources

While the bioinformatics approach is robust, it would benefit from a more detailed description of the methods and algorithms used. Additionally, the selection criteria for the breast cancer data should be justified more thoroughly to avoid selection bias.

Statistical Analysis

The statistical methods used need more comprehensive detailing. For complex analyses, ensure the statistical assumptions and any transformations of data are clearly explained. Include more information on the statistical tests used for hypothesis testing and the justification for their use.

Comparative Analysis

The comparison between hereditary breast cancers and those potentially caused by EBV is insightful. However, a more detailed comparative analysis would strengthen the argument. This could include molecular or genetic profiling comparisons.

Discussion on Contradictory or Supporting Evidence

The discussion section should address not only the supporting evidence but also any contradictory findings in the literature. This balance is crucial for a nuanced understanding of the subject.

Implications and Future Research Directions

The implications of these findings are profound but need clearer articulation. Discuss the potential impact on breast cancer treatment and prevention strategies. Also, outline future research directions, particularly in clinical or experimental studies, to confirm these bioinformatics findings.

References

Please add more background information about breast cancer (please cite: 1. Cao Y, Efetov S, He M, et al. Updated clinical perspectives and challenges of chimeric antigen receptor-T cell therapy in colorectal cancer and invasive breast cancer. Arch Immunol Ther Exp (Warsz). Aug 11, 2023;71(1):19. [doi: 10.1007/s00005-023-00684-x] [Medline: 37566162]; and 2. Liu Y, Lu S, Sun Y, et al. Deciphering the role of QPCTL in glioma progression and cancer immunotherapy. Front Immunol.

Mar 29, 2023;14:1166377. [doi: 10.3389/fimmu.2023.1166377] [Medline: 37063864]).

Concluding Remarks

The paper presents a novel and potentially significant hypothesis linking EBV to breast cancer. However, it requires major revisions to enhance its methodological rigor, clarity, and comprehensiveness. Addressing these concerns will significantly strengthen the manuscript's impact and contribution to the field.

Conflicts of Interest

None declared.

Reference

1. Friedenson B. Identifying safeguards disabled by Epstein-Barr virus infections in genomes from patients with breast cancer: chromosomal bioinformatics analysis. JMIRx Med 2025;6:e50712. [doi: 10.2196/50712]

Abbreviations

EBV: Epstein-Barr virus

Edited by A Schwartz; submitted 13.12.24; this is a non–peer-reviewed article;accepted 13.12.24; published 29.01.25.

Please cite as:
Anonymous
Peer Review of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis"
JMIRx Med 2025;6:e70039
URL: https://xmed.jmir.org/2025/1/e70039
doi:10.2196/70039

Peer Review of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection"

Reenu Singh

Indian Institute of Management Mumbai, Mumbai, India

Related Articles:

Companion article: https://arxiv.org/abs/2410.17459v1

Companion article: https://med.jmirx.org/2025/1/e72527

Companion article: https://med.jmirx.org/2025/1/e70100

(JMIRx Med 2025;6:e72523) doi:10.2196/72523

KEYWORDS

privacy-preserving AI; latent space projection; data obfuscation; AI governance; machine learning privacy; differential privacy; k-anonymity; HIPAA; GDPR; compliance; data utility; privacy-utility trade-off; responsible AI; medical imaging privacy; secure data sharing; LSP; artificial intelligence

This is a peer-review report for "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection."

Round 1 Review

Specific Comments

Major Comments

1. What was the basis of taking up health care cancer diagnosis and financial fraud for the study [1]? Will latent space projection be an effective method for privacy protection in speech therapy to analyze audio datasets to assist in diagnosing and treating speech-related disorders; in medical imaging video datasets from endoscopy, ultrasounds, and robotic surgeries for diagnostics and artificial intelligence–assisted tools; and in telemedicine to analyze video feeds for remote consultations and diagnoses?
2. The basic structure of the paper is missing. Please follow the guidelines of journal paper writing with distinctly visible sections of Introduction, Method, Result/Findings, Discussion, and Limitations with future scope and conclusion. The introduction, background, and related work should be written cohesively, and all should come under the Introduction heading.
3. The statistical tables are in excess. The tables and values should be talked about in written form. Limit the number of images and tables to 5 - 6 or according to the journal guidelines. Use an appendix for the flowchart and any other tabular data that is too lengthy.
4. Explanations of tables and figures should be in paragraph form. Please cite literature where comparative inference and process-specific benefits and drawbacks are mentioned. Examples are Tables 1-5. For writing sections like "Comparative Analysis with Existing Techniques," all the subparts should be written in paragraphs and discuss the values and analysis only, and put them in their respective paragraphs, removing the tabular data. Please use appendices for excessive tables. Within the body of the research paper, 5 - 6 figures and tables are sufficient; the rest should be put in appendices.
5. In "Latency and Performance analysis, part A" and "Performance optimization" are mentions of the literature, which should be present as part of the literature in the Introduction paragraph. Restating the literature again is redundant. Stick to the structure of the journal paper. Please cite references to support the claims, such as "real-time requirements of financial systems" under the section of Real-Time Performance.
6. "Scalability analysis" and other sections: What were the criteria for the choice of datasets for the study for the case studies? What were the data sizes? Give specifications in the first paragraph of respective case studies. Presenting the details about the process of procurement of files, data extraction, limitations in data handling, etc. Are there any limitations in adopting the latent space projection methods?

Round 2 Review

General Comments

This paper is highly relevant to health care, particularly in the context of privacy management of data during the analysis of imagery.

Specific Comments

Major Comments

1. The case studies should be written in a more descriptive style. Please reduce the use of numbered or bullet points (in the Introduction, Method, and Result) to align with the formal writing style typically suitable for journal papers.

2. Please rephrase the description of Table 3 (immediately following the table) in a narrative style. This approach enhances the readability of the article.

3. Two figures should not be positioned consecutively. Include some text between Figure 3 and Figure 4. Adjust and reorganize the content to ensure a smooth flow.

Minor Comments

4. The titles of tables and figures should be presented as captions. Revise the captions to ensure they do not begin with a verb.

Conflicts of Interest

None declared.

Reference

1. Vaijainthymala Krishnamoorthy M. Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection. JMIRx Med 2025;6:e70100. [doi: 10.2196/70100]

Edited by CN Hang; submitted 11.02.25; this is a non–peer-reviewed article;accepted 11.02.25; published 12.03.25.

Please cite as:
Singh R
Peer Review of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection"
JMIRx Med 2025;6:e72523
URL: https://xmed.jmir.org/2025/1/e72523
doi:10.2196/72523

Peer Review of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis"

Anonymous

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.03.23292185v1

Companion article: https://med.jmirx.org/2025/1/e69307

Companion article: https://med.jmirx.org/2025/1/e50712

(JMIRx Med 2025;6:e70041) doi:10.2196/70041

KEYWORDS

breast; cancer; oncology; ovarian; virus; viral; Epstein-Barr; herpes; bioinformatics; chromosome; gene; genetic; chromosomal; DNA; genomic; BRCA; metastasis; biology

This is the peer-review report for "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis."

Round 1 Review

Dear Author,

After a thorough review of the paper titled "Herpesvirus infections eliminate safeguards against breast cancer and its metastasis: comparable to hereditary breast cancers" [1] by Bernard Friedenson, here is the negative feedback and evaluation, along with a recommendation for the inclusion of a specific article in the discussion section.

Negative Feedback and Evaluation

Clarity and Scope

The paper ambitiously attempts to link Epstein-Barr virus (EBV) infections to breast cancer development and metastasis. While the hypothesis is intriguing, the narrative sometimes lacks clarity and could benefit from a more focused scope. The vast amount of data and the complex mechanisms presented can be overwhelming and occasionally detract from the main message.

Methodological Concerns

The reliance on bioinformatics analyses and previously published datasets raises questions about the direct experimental validation of the proposed mechanisms. Although the computational approach is valid, the absence of direct experimental evidence or validation in breast cancer samples limits the strength of the conclusions.

Interpretation of Data

The interpretation of viral homology and its impact on cancer development is speculative in several sections. The connections made between EBV infections, chromosomal breakpoints, and cancerous mutations rely heavily on correlative data without sufficient causal evidence. A more cautious interpretation of the results, highlighting the need for further experimental validation, would strengthen the manuscript.

Consideration of Alternate Hypotheses

The paper could benefit from a more balanced discussion of alternative hypotheses explaining the observed data. For instance, the role of other environmental, genetic, or lifestyle factors in breast cancer development is not adequately considered. Acknowledging and discussing these potential confounders would provide a more comprehensive understanding of the complex etiology of breast cancer.

References and Current Literature

While the paper cites a significant amount of relevant literature, it sometimes overlooks recent studies that could either support or challenge the proposed hypotheses. Incorporating a more current and diverse range of references would enhance the paper's relevance and credibility.

Recommendation for Discussion Inclusion

To broaden the discussion and contextualize the findings within the broader research landscape, it is recommended to include the following article in the discussion section.

Al-Awaida W, Al-Ameer HJ, Sharab A, Akasheh RT. Modulation of wheatgrass (*Triticum aestivum Linn*) toxicity against breast cancer cell lines by simulated microgravity. Curr Res Toxicol. Sep 19, 2023;5:100127. [doi: 10.1016/j.crtox.2023.100127] [Medline: 37767028]

Incorporating this article could provide valuable insights into innovative approaches for studying cancer therapies. Specifically, the effects of simulated microgravity on the efficacy of natural compounds like wheatgrass against breast

cancer could open up new avenues for research on the environmental and physical conditions affecting cancer treatment outcomes. Discussing this study would enrich the manuscript by introducing the concept of microgravity as a novel factor influencing cancer cell behavior and therapy resistance, thereby offering a broader perspective on cancer research methodologies and therapeutic strategies.

Round 2 Review

General Comments

This paper tests the idea that EBV infections can help cause breast cancer by weakening the body's defenses against cancer. The study uses bioinformatics to compare chromosome breakpoints in breast cancer to those in cancers known to be caused by EBV. The results show that EBV might play a role in breast cancer by damaging important cell functions.

Specific Comments

Major Comments

The methods section needs more details about how the datasets were chosen and combined.

The discussion should explain more about how EBV might cause the chromosome breaks and rearrangements seen in breast cancer.

More data or references are needed to support the idea that EBV helps breast cancer spread to other parts of the body.

Minor Comments

Adding more references would strengthen the sections that talk about how EBV affects breast cancer.

Figures and tables should be clearly mentioned in the text to help readers follow the data.

Some parts of the manuscript need clearer writing and better organization, especially where complex bioinformatics results are explained.

The abstract should be revised to clearly highlight the main findings and why they are important.

Make sure all abbreviations are defined when they are first used to help readers understand the text better.

Conflicts of Interest

None declared.

Reference

1. Friedenson B. Identifying safeguards disabled by Epstein-Barr virus infections in genomes from patients with breast cancer: chromosomal bioinformatics analysis. JMIRx Med 2025;6:e50712. [doi: 10.2196/50712]

Abbreviations

EBV: Epstein-Barr virus

Edited by A Schwartz; submitted 13.12.24; this is a non–peer-reviewed article;accepted 13.12.24; published 29.01.25.

Please cite as:
Anonymous
Peer Review of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis"
JMIRx Med 2025;6:e70041
URL: https://xmed.jmir.org/2025/1/e70041
doi:10.2196/70041

Peer Review of “Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection”

Trutz Bommhardt

University of Wuppertal, Wuppertal, Germany

Related Articles:

Companion article: https://arxiv.org/abs/2410.17459v1

Companion article: https://med.jmirx.org/2025/1/e72527

Companion article: https://med.jmirx.org/2025/1/e70100

(JMIRx Med 2025;6:e72525) doi:10.2196/72525

KEYWORDS

privacy-preserving AI; latent space projection; data obfuscation; AI Governance; machine learning privacy; differential privacy; k-anonymity; HIPAA; GDPR; compliance; data utility; privacy-utility trade-off; responsible AI; medical imaging privacy; secure data sharing; LSP; artificial intelligence

This is a peer-review report for “Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection.”

Round 1 Review

General Comments

I thoroughly enjoyed reading this paper [1] as it is a well-written article that will make an important contribution to the literature on the development of privacy-preserving artificial intelligence (AI) governance. I have attached a few comments to improve the study.

Specific Comments

Major Comments

Something like a discussion that embeds the latent space projection for AI governance and the results in the current scientific debate is missing before or after Chapter VII.

Minor Comments

In Chapter II B (Existing privacy-preserving techniques), please provide some further sources to demonstrate that the challenges mentioned are still relevant, as some sources are relatively old (eg, from 2009).

Conflicts of Interest

None declared.

Reference

1. Vaijainthymala Krishnamoorthy M. Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection. JMIRx Med 2025;6:e70100. [doi: 10.2196/70100]

Abbreviations

AI: artificial intelligence

Edited by CN Hang; submitted 11.02.25; this is a non–peer-reviewed article;accepted 11.02.25; published 12.03.25.

Please cite as:
Bommhardt T
Peer Review of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection"
JMIRx Med 2025;6:e72525
URL: https://xmed.jmir.org/2025/1/e72525
doi:10.2196/72525

Authors' Response to Peer Reviews of "Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review"

Feryal Kurdi*, MD; Yahya Kurdi*, MD; Igor Vladimirovich Reshetov, MD, PhD

Department of Oncology, Radiotherapy and Plastic and Reconstructive Surgery, Sechenov University, Bolshaya Pirogovskaya, 6c1, Moscow, Russian Federation

*these authors contributed equally

Corresponding Author:
Feryal Kurdi, MD
Department of Oncology, Radiotherapy and Plastic and Reconstructive Surgery, Sechenov University, Bolshaya Pirogovskaya, 6c1, Moscow, Russian Federation

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.07.30.24311256v1

Companion article: https://med.jmirx.org/2025/1/e69705

Companion article: https://med.jmirx.org/2025/1/e66213

(JMIRx Med 2025;6:e68769) doi:10.2196/68769

KEYWORDS

indocyanine green; ICG; sentinel lymph node; breast cancer; breast; fluorescence; axillary lymph node mapping; NIR; surgical planning; near-infrared

This is the authors' response to peer-review reports for "Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review."

Round 1 Review

Anonymous [1]

General Comments

This paper [2] summarized the application value and existing problems of indocyanine green (ICG) in sentinel lymph node (SLN) biopsy of early breast cancer, which has positive significance for improving the accuracy of clinical SLN detection. This study has certain clinical value.

Response: Thank you for your thoughtful comments and feedback on our paper. Below are my responses to your points.

Specific Comments

Major Comments

1. Due to the high hardware requirements for the clinical application of ICG, the number of relevant studies in the search is relatively small. It is hoped that the author can search the recent, relevant literature to improve the credibility of this review.

Response: This paper is a protocol for a scoping review, serving as a roadmap for the search strategy and inclusion criteria that we will follow. As such, it outlines our plan rather than reporting the outcomes of the literature search. As noted in Multimedia Appendix 1, we will conduct a comprehensive search across multiple databases to ensure the inclusion of all relevant, recent studies.

2. It is hoped that the author will analyze and compare the advantages and disadvantages of ICG and traditional SLN biopsy methods, so as to guide clinicians to adopt appropriate methods for appropriate patients.

Response: As indicated in Multimedia Appendix 1, this comparison is a core objective of our review. We hope these clarifications address your concerns.

References

1. Anonymous. Peer review of "Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review. JMIRx Med 2024;5:e69705. [doi: 10.2196/69705]

2. Kurdi F, Kurdi Y, Reshetov IV. Applications of indocyanine green in breast cancer for sentinel lymph node mapping: protocol for a scoping review. JMIRx Med 2024;5:e66213. [doi: 10.2196/66213]

Abbreviations

ICG: indocyanine green
SLN: sentinel lymph node

Edited by S Tungjitviboonkun; submitted 13.11.24; this is a non–peer-reviewed article;accepted 13.11.24; published 06.01.25.

Please cite as:
Kurdi F, Kurdi Y, Reshetov IV
Authors' Response to Peer Reviews of "Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review"
JMIRx Med 2025;6:e68769
URL: https://xmed.jmir.org/2025/1/e68769
doi:10.2196/68769

Author's Response to Peer Reviews of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis"

Bernard Friedenson, PhD

Department of Biochemistry and Medical Genetics, Cancer Center, University of Illinois Chicago, 900 s Ashland, Chicago, IL, United States

Corresponding Author:
Bernard Friedenson, PhD
Department of Biochemistry and Medical Genetics, Cancer Center, University of Illinois Chicago, 900 s Ashland, Chicago, IL, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.03.23292185v1

Companion article: https://med.jmirx.org/2025/1/e70039

Companion article: https://med.jmirx.org/2025/1/e70041

Companion article: https://med.jmirx.org/2025/1/e50712

(JMIRx Med 2025;6:e69307) doi:10.2196/69307

KEYWORDS

breast; cancer; oncology; ovarian; virus; viral; Epstein-Barr; herpes; bioinformatics; chromosome; gene; genetic; chromosomal; DNA; genomic; BRCA; metastasis; biology

This is the author's response to peer-review reports for "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis."

Round 1 Review

Anonymous [1]

Review Report With Major Revisions for the Paper

Title: "Herpesvirus infections eliminate safeguards against breast cancer and its metastasis: comparable to hereditary breast cancers"

Summary

The paper [2] hypothesizes that Epstein-Barr virus (EBV) infections promote breast cancer by disabling cancer safeguards. It is a bioinformatics analysis of public information from about 2100 breast cancers. The study finds that breast and ovarian cancer breakpoints cluster around EBV-associated cancer breakpoints, suggesting a significant role of EBV in promoting these cancers. The paper also identifies similarities in the molecular and cellular disruptions caused by EBV with those found in hereditary breast cancers.

Major Revisions Needed

Clarification of Hypotheses and Objectives

The hypothesis, while intriguing, needs clearer articulation. Specifically, the connection between EBV and breast cancer needs more explicit theoretical underpinning. Clarify the objectives and expected outcomes of the study at the outset.

Response: The objectives and expected outcomes of the study were clarified at the outset in the Abstract and Introduction.

Methodological Rigor and Data Sources

While the bioinformatics approach is robust, it would benefit from a more detailed description of the methods and algorithms used. Additionally, the selection criteria for the breast cancer data should be justified more thoroughly to avoid selection bias.

Response: A more detailed description of the methods and algorithms used has been added in the Methods section (page 6).

Statistical Analysis

The statistical methods used need more comprehensive detailing. For complex analyses, ensure the statistical assumptions and any transformations of data are clearly explained. Include more information on the statistical tests used for hypothesis testing and the justification for their use.

Response: I included more information on the statistical tests, the justification, and limitations of their use (page 7).

Comparative Analysis

The comparison between hereditary breast cancers and those potentially caused by EBV is insightful. However, a more detailed comparative analysis would strengthen the argument. This could include molecular or genetic profiling comparisons.

Response: I added a more detailed comparative analysis with results in Figure 2H and Table S2, as described on page 10.

Discussion on Contradictory or Supporting Evidence

The discussion section should address not only the supporting evidence but also any contradictory findings in the literature. This balance is crucial for a nuanced understanding of the subject.

Response: The paper's hypothesis more clearly accounts for the absence of demonstrable EBV infection in breast cancer, explaining contradictory results. The other contradictory result posits an imperfect palindrome on chromosome 11. This result is tested on page 13.

Implications and Future Research Directions

The implications of these findings are profound but need clearer articulation. Discuss the potential impact on breast cancer treatment and prevention strategies. Also, outline future research directions, particularly in clinical or experimental studies to confirm these bioinformatics findings.

Response: I articulated the implications of these finding more clearly with their impact on breast cancer treatment and prevention strategies. I also outlined future research directions with clinical or experimental studies to confirm the bioinformatics findings (Discussion, page 16).

References

Please add more background information about breast cancer (please cite: 1. Cao Y, Efetov S, He M, et al. Updated clinical perspectives and challenges of chimeric antigen receptor-T cell therapy in colorectal cancer and invasive breast cancer. Arch Immunol Ther Exp (Warsz). Aug 11, 2023;71(1):19. [doi: 10.1007/s00005-023-00684-x] [Medline: 37566162]; and 2. Liu Y, Lu S, Sun Y, et al. Deciphering the role of QPCTL in glioma progression and cancer immunotherapy. Front Immunol. Mar 29, 2023;14:1166377. [doi: 10.3389/fimmu.2023.1166377] [Medline: 37063864]).

Response: I added these references.

Concluding Remarks

The paper presents a novel and potentially significant hypothesis linking EBV to breast cancer. However, it requires major revisions to enhance its methodological rigor, clarity, and comprehensiveness. Addressing these concerns will significantly strengthen the manuscript's impact and contribution to the field.

Anonymous [3]

Dear Author,

After a thorough review of the paper titled "Herpesvirus infections eliminate safeguards against breast cancer and its metastasis: comparable to hereditary breast cancers" by Bernard Friedenson, here is the negative feedback and evaluation, along with a recommendation for the inclusion of a specific article in the discussion section.

Negative Feedback and Evaluation

Clarity and Scope

The paper ambitiously attempts to link Epstein-Barr virus (EBV) infections to breast cancer development and metastasis. While the hypothesis is intriguing, the narrative sometimes lacks clarity and could benefit from a more focused scope. The vast amount of data and the complex mechanisms presented can be overwhelming and occasionally detract from the main message.

Response: I focused the scope in this revision in the Abstract and Introduction.

Methodological Concerns

The reliance on bioinformatics analyses and previously published datasets raises questions about the direct experimental validation of the proposed mechanisms. Although the computational approach is valid, the absence of direct experimental evidence or validation in breast cancer samples limits the strength of the conclusions.

Response: I explained in the Discussion section that direct experimental evidence or validation has already been done. EBV-infected human mammary epithelial cells produce breast cancer in immunosuppressed mice (page 17).

Interpretation of Data

The interpretation of viral homology and its impact on cancer development is speculative in several sections. The connections made between EBV infections, chromosomal breakpoints, and cancerous mutations rely heavily on correlative data without sufficient causal evidence. A more cautious interpretation of the results, highlighting the need for further experimental validation, would strengthen the manuscript.

Response: I added more evidence (Figure 2H and Table S2) to the association of EBV infection and cancer development and took greater care throughout to interpret the results more cautiously.

Consideration of Alternate Hypotheses

The paper could benefit from a more balanced discussion of alternative hypotheses explaining the observed data. For instance, the role of other environmental, genetic, or lifestyle factors in breast cancer development is not adequately considered. Acknowledging and discussing these potential confounders would provide a more comprehensive understanding of the complex etiology of breast cancer.

Response: I explained how EBV relates to alternate hypotheses and exacerbates the effects of other known breast cancer risk factors (page 16).

References and Current Literature

While the paper cites a significant amount of relevant literature, it sometimes overlooks recent studies that could either support or challenge the proposed hypotheses. Incorporating a more

current and diverse range of references would enhance the paper's relevance and credibility.

Response: I included more information from more current and diverse ranges of references.

Recommendation for Discussion Inclusion

To broaden the discussion and contextualize the findings within the broader research landscape, it is recommended to include the following article in the discussion section.

Al-Awaida W, Al-Ameer HJ, Sharab A, Akasheh RT. Modulation of wheatgrass (Triticum aestivum Linn) toxicity against breast cancer cell lines by simulated microgravity. Curr Res Toxicol. Sep 19, 2023;5:100127. [doi: 10.1016/j.crtox.2023.100127] [Medline: 37767028]

Incorporating this article could provide valuable insights into innovative approaches for studying cancer therapies. Specifically, the effects of simulated microgravity on the efficacy of natural compounds like wheatgrass against breast cancer could open up new avenues for research on the environmental and physical conditions affecting cancer treatment outcomes. Discussing this study would enrich the manuscript by introducing the concept of microgravity as a novel factor influencing cancer cell behavior and therapy resistance, thereby offering a broader perspective on cancer research methodologies and therapeutic strategies.

Response: I could not find a way to apply and cite this interesting work since it was so far afield from the manuscript.

Round 2 Review

Anonymous [3]

General Comments

This paper tests the idea that EBV infections can help cause breast cancer by weakening the body's defenses against cancer. The study uses bioinformatics to compare chromosome breakpoints in breast cancer to those in cancers known to be caused by EBV. The results show that EBV might play a role in breast cancer by damaging important cell functions.

Specific Comments

Major Comments

The methods section needs more details about how the datasets were chosen and combined.

Response: More details on how the datasets were chosen have been added.

The discussion should explain more about how EBV might cause the chromosome breaks and rearrangements seen in breast cancer.

Response: The discussion includes an expanded explanation about how EBV might cause the chromosome breaks and rearrangements seen in breast cancer.

More data or references are needed to support the idea that EBV helps breast cancer spread to other parts of the body.

Response: A new Figure 7 and more data have been added. Additional references have also been added, and the metastasis topic has been clarified and expanded.

Minor Comments

Adding more references would strengthen the sections that talk about how EBV affects breast cancer.

Response: Many more references have been added.

Figures and tables should be clearly mentioned in the text to help readers follow the data.

Response: Figures and tables are now more prominently mentioned in the text.

Some parts of the manuscript need clearer writing and better organization, especially where complex bioinformatics results are explained.

Response: I revised the manuscript with clearer writing and better organization, especially where complex bioinformatics results are explained.

The abstract should be revised to clearly highlight the main findings and why they are important.

Response: I revised the Abstract to highlight the main findings and why they are important.

Make sure all abbreviations are defined when they are first used to help readers understand the text better.

Response: I went through the manuscript to be sure all abbreviations were defined. I also added a glossary containing abbreviations, gene names, and viruses.

References

1. Anonymous. Peer review of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis". JMIRx Med 2025;6:e70039. [doi: 10.2196/70039]
2. Friedenson B. Identifying safeguards disabled by Epstein-Barr virus infections in genomes from patients with breast cancer: chromosomal bioinformatics analysis. JMIRx Med 2025;6:e50712. [doi: 10.2196/50712]
3. Anonymous. Peer review of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis". JMIRx Med 2025;6:e70041. [doi: 10.2196/70041]

Abbreviations

EBV: Epstein-Barr virus

Edited by A Schwartz; submitted 26.11.24; this is a non–peer-reviewed article;accepted 26.11.24; published 29.01.25.

Please cite as:
Friedenson B
Author's Response to Peer Reviews of "Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis"
JMIRx Med 2025;6:e69307
URL: https://xmed.jmir.org/2025/1/e69307
doi:10.2196/69307

Authors' Response to Peer Reviews of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection"

Mahesh Vaijainthymala Krishnamoorthy, BE

Stelmith, LLC, 2333 Aberdeen Pl, Carollton, TX, United States

Corresponding Author:
Mahesh Vaijainthymala Krishnamoorthy, BE
Stelmith, LLC, 2333 Aberdeen Pl, Carollton, TX, United States

Related Articles:

Companion article: https://arxiv.org/abs/2410.17459v1

Companion article: https://med.jmirx.org/2025/1/e72523

Companion article: https://med.jmirx.org/2025/1/e72525

Companion article: https://med.jmirx.org/2025/1/e70100

(JMIRx Med 2025;6:e72527) doi:10.2196/72527

KEYWORDS

privacy-preserving AI; latent space projection; data obfuscation; AI governance; machine learning privacy; differential privacy; k-anonymity; HIPAA; GDPR; compliance; data utility; privacy-utility trade-off; responsible AI; medical imaging privacy; secure data sharing; LSP; artificial intelligence

This is the authors' response to peer-review reports for "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection."

Round 1 Review

Reviewer AP [1]

Specific Comments

Major Comments

1. What was the basis of taking up health care cancer diagnosis and financial fraud for the study [2]? Will latent space projection be an effective method for privacy protection in speech therapy to analyze audio datasets to assist in diagnosing and treating speech-related disorders; in medical imaging video datasets from endoscopy, ultrasounds, and robotic surgeries for diagnostics and artificial intelligence (AI)–assisted tools; and in telemedicine to analyze video feeds for remote consultations and diagnoses?

Response: The basis for taking this up is to show data privacy through images and records for individuals. I would love to extend the research and will work on another paper for your suggestions. Thanks for the suggestion.

2. The basic structure of the paper is missing. Please follow the guidelines of journal paper writing with distinctly visible sections of Introduction, Method, Result/Findings, Discussion, and Limitations with future scope and conclusion. The introduction, background, and related work should be written cohesively, and all should come under the Introduction heading.

Response: I have revised the paper with major formatting changes and made it follow the Introduction-Methods-Results-Discussion formatting style as per the suggestion.

3. The statistical tables are in excess. The tables and values should be talked about in written form. Limit the number of images and tables to 5‐6 or according to the journal guidelines. Use an appendix for the flowchart and any other tabular data that is too lengthy.

Response: Statistical tables were reduced to only 3, and Figures are limited to 6 in total, but the flowchart is necessary inside the main paper.

4. Explanations of tables and figures should be in paragraph form. Please cite literature where comparative inference and process-specific benefits and drawbacks are mentioned. Examples are Tables 1-5. For writing sections like "Comparative Analysis with Existing Techniques," all the subparts should be written in paragraphs and discuss the values and analysis only, and put them in their respective paragraphs, removing the tabular data. Please use appendices for excessive tables. Within the body of the research paper, 5‐6 figures and tables are sufficient; the rest should be put in appendices.

Response: Tables have been removed and converted into paragraphs

5. In "Latency and Performance analysis, part A" and "Performance optimization" are mentions of the literature, which should be present as part of the literature in the Introduction paragraph. Restating the literature again is redundant. Stick to the structure of the journal paper. Please cite references to support the claims, such as "real-time requirements of financial systems" under the section of Real-Time Performance.

Response: Thanks; moved to the Literature section and removed from there.

6. "Scalability analysis" and other sections: What were the criteria for the choice of datasets for the study for the case studies? What were the data sizes? Give specifications in the first paragraph of respective case studies. Presenting the details about the process of procurement of files, data extraction, limitations in data handling, etc. Are there any limitations in adopting the latent space projection methods?

Response: Scalability analysis was added with the source of the dataset and the data extraction and limitations. Mostly, there are a lot of advantages compared to other privacy-preserving techniques in latent space projection; the comparative analysis proves that, and a few limitations were added as well.

Reviewer AR [3]

General Comments

I thoroughly enjoyed reading this paper as it is a well-written article that will make an important contribution to the literature on the development of privacy-preserving AI governance. I have attached a few comments to improve the study.

Response: Thanks for the compliment. Thanks for your time.

Specific Comments

Major Comments

Something like a discussion that embeds the latent space projection for AI governance and the results in the current scientific debate is missing before or after Chapter VII.

Minor Comments

In Chapter II B (Existing privacy-preserving techniques), please provide some further sources to demonstrate that the challenges mentioned are still relevant, as some sources are relatively old (eg, from 2009).

Response: I tried to address all your comments.

Round 2 Review

Reviewer AP

General Comments

This paper is highly relevant to health care, particularly in the context of privacy management of data during the analysis of imagery.

Response: Thanks for your time and effort. I appreciate it. Your comments were valuable. I addressed all your comments in this revision.

Specific Comments

Major Comments

1. The case studies should be written in a more descriptive style. Please reduce the use of numbered or bullet points (in the Introduction, Method, and Result) to align with the formal writing style typically suitable for journal papers.

Response: Removed all the bullets and converted most of them into paragraphs; some were aligned as paragraphs, but the bullet and numbered points were removed. The paper is in the Introduction-Methods-Results-Discussion format.

2. Please rephrase the description of Table 3 (immediately following the table) in a narrative style. This approach enhances the readability of the article.

Response: Rephrased the description for all the tables and figures, added descriptions for two other figures, explaining the figures deeply to make it more even, uniform, and readable, and for smooth flow.

3. Two figures should not be positioned consecutively. Include some text between Figure 3 and Figure 4. Adjust and reorganize the content to ensure a smooth flow.

Response: Addressed by adding content between 2 figures; now it makes it more readable and flows smoothly. Thanks.

Minor Comments

4. The titles of tables and figures should be presented as captions. Revise the captions to ensure they do not begin with a verb.

Response: Revised all the captions for tables and figures and made them capitalized and more readable.

Thanks for your comments.

References

1. Singh R. Peer review of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection". JMIRx Med 2025;6:e72523. [doi: 10.2196/72523]
2. Vaijainthymala Krishnamoorthy M. Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection. JMIRx Med 2025;6:e70100. [doi: 10.2196/70100]
3. Bommhardt T. Peer review of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection". JMIRx Med 2025;6:e72525. [doi: 10.2196/72525]

Abbreviations

AI: artificial intelligence

Edited by CN Hang; submitted 11.02.25; this is a non–peer-reviewed article;accepted 11.02.25; published 12.03.25.

Please cite as:
Vaijainthymala Krishnamoorthy M
Authors' Response to Peer Reviews of "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection"
JMIRx Med 2025;6:e72527
URL: https://xmed.jmir.org/2025/1/e72527
doi:10.2196/72527

Authors' Response to Peer Reviews of "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study"

Ajit Kerketta*, MHA; Raghavendra A N*, PhD

CHRIST (Deemed to be University), Hosur Road, Bhavani Nagar, Bengaluru, India

*all authors contributed equally

Corresponding Author:
Ajit Kerketta, MHA
CHRIST (Deemed to be University), Hosur Road, Bhavani Nagar, Bengaluru, India

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.04.12.23288461v1

Companion article: https://med.jmirx.org/2025/1/e70808

Companion article: https://med.jmirx.org/2025/1/e48346

(JMIRx Med 2025;6:e70059) doi:10.2196/70059

KEYWORDS

rural alimentation; community health workers; motivation; retention; rural health; rural nutrition; workforce

This is the authors' response to peer-review reports for "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study."

Round 1 Review [1]

General Comments

This paper [2] has given the impression that the researcher has done thorough homework before starting the research and it is evident in the paper. Case methodology and thematic analysis are a few of the approaches that depict the quality of the paper. Overall, as a reviewer, it is my opinion that the research paper is of quality.

Specific Comments

1. A few more factors like government initiatives should be included in studying the impact on the motivation and retention of community health workers.

Response: Factors such as government initiatives and policies have been additionally incorporated into the Discussion section.

Major Comments

1. I feel that the analysis also can include education as a parameter.

2. The thematic analysis is one of the strengths of this research and is appreciated.

Response: Due to time constraints, education could not be included as a sample parameter.

Minor Comments

1. Common wording should be used in every section of the paper, like qualitative case research methodology and qualitative case research.

Response: The term "qualitative case research" has been consistently used throughout the study.

References

1. Kumar Thalari S. Peer review of "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study". JMIRx Med 2025;6:e70808. [doi: 10.2196/70808]
2. Kerketta A, A N R. The impact of rural alimentation on the motivation and retention of Indigenous community health workers in India: qualitative study. JMIRx Med 2025;6:e48346. [doi: 10.2196/48346]

Edited by A Schwartz; submitted 13.12.24; this is a non–peer-reviewed article;accepted 13.12.24; published 23.01.25.

Please cite as:
Kerketta A, A N R
Authors' Response to Peer Reviews of "The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: Qualitative Study"
JMIRx Med 2025;6:e70059
URL: https://xmed.jmir.org/2025/1/e70059
doi:10.2196/70059

Authors' Response to Peer Reviews of "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study"

Tahazid Tamannur[1], MPH; Sadhan Kumar Das[1], MPH; Arifatun Nesa[2], MPH; Foijun Nahar[1], MPH; Nadia Nowshin[1], MPH; Tasnim Haque Binty[1], MPH; Shafiul Azam Shakil[2], MPH; Shuvojit Kumar Kundu[3], MPH; Md Abu Bakkar Siddik[4], MPH; Shafkat Mahmud Rafsun[5], MPH; Umme Habiba[6], MPH; Zaki Farhana[7], MS; Hafiza Sultana[1], MPhil; Anton Abdulbasah Kamil[8], PhD; Mohammad Meshbahur Rahman[9], MS

[1]Department of Health Education, National Institute of Preventive and Social Medicine, Dhaka, Bangladesh
[2]Department of Public Health and Hospital Administration, National Institute of Preventive and Social Medicine, Mohakhali, Dhaka, Bangladesh
[3]Directorate General of Health Services, Ministry of Health & Family Welfare, Government of the People's Republic of Bangladesh, Dhaka, Bangladesh
[4]School of the Environment, Nanjing University, Nanjing, China
[5]Dental Speciality Center, Dhaka, Bangladesh
[6]BRAC James P Grant School of Public Health, BRAC University, Dhaka, Bangladesh
[7]Credit Information Bureau, Bangladesh Bank, Dhaka, Bangladesh
[8]Department of Business Administration, Istanbul Gelisim University, Istanbul, Turkey
[9]Department of Biostatistics, National Institute of Preventive and Social Medicine, Dhaka, Bangladesh

Corresponding Author:
Mohammad Meshbahur Rahman, MS
Department of Biostatistics, National Institute of Preventive and Social Medicine, Dhaka, Bangladesh

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.04.05.24305403v1

Companion article: https://med.jmirx.org/2025/1/e70142

Companion article: https://med.jmirx.org/2025/1/e70144

Companion article: https://med.jmirx.org/2025/1/e59379

(JMIRx Med 2025;6:e70145) doi:10.2196/70145

KEYWORDS

mothers' knowledge and practices; oral hygiene; child oral health; Bangladesh

This is the authors' response to peer-review reports for "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study."

Round 1 Review

Reviewer BZ [1]

This is an interesting piece of research [2], which highlights mothers' knowledge and practices regarding their children's oral health in Dhaka City. However, several issues made the study scientifically questionable. The major issues are as follows. The study included mothers from two hospitals in Dhaka City, but the title of the study does not mention this. The sample selection from the mothers visiting the hospitals might not represent general mothers from the whole of Dhaka. Thus, this study might not be generalizable to all mothers in Dhaka City.

Response: The authors are grateful to the reviewers for critically reviewing our manuscript. We agree with the comments. Respondents of this study were the mothers visiting the tertiary-level hospitals of Dhaka City. Generally, the respondents visiting hospitals belonged to all administrative wards (small regions of Dhaka), and it is convenient to get the mothers with children aged 5 - 9 years to interview. That is why we chose tertiary-level hospitals to reach the respondents. However, we revised our manuscript title and omitted "Dhaka" from the title. The new title is "Knowledge and practices towards oral hygiene of children aged 5 - 9 years old: a cross-sectional study among mothers visited tertiary level hospitals."

Introduction

Revise the last paragraph of the Introduction to highlight the study gap in Bangladesh and clearly state the objective of the

study. Use the formal word "mother" and avoid the word "moms."

Response: We appreciate the reviewer for this comment. We revised the Introduction of our study and replaced the word "Moms" with mother.

Methods

Study Setting and Participants

Give clear reasoning as to why you selected study participants from the hospitals. The last line is confusing. It is not clear whether the participants filled out the questionnaire on their own or they were interviewed by the enumerators.

Response: We are thankful to the reviewer for this comment. Respondents of this study were the mothers visiting the tertiary-level hospitals of Dhaka City. Generally, the respondents who visited hospitals belonged to all administrative wards (small regions of Dhaka), and it was convenient to get this group of mothers with children aged 5 - 9 years to interview. That is why we chose tertiary-level hospitals to reach the respondents. However, we revised our manuscript title and omitted "Dhaka" from the title. We interviewed the respondents, and the sentence was revised in our revised manuscript.

Sampling Technique

Please mention the nonresponse bias for the convenient sampling. Give a short description of the pretesting mentioning the number of samples, period, and location for it.

Response: We are again thankful to the reviewer. While we had a 5% nonresponse rate in our final survey, we found less than 5% (2 of 50 mothers refused to be involved in the study) as the nonresponse rate during pretesting of our study. The description of the pretest has been given in our revised manuscript. In our main survey, the nonresponse rate was 2%.

Measurement of Knowledge and Practice Score

Give the 15 knowledge-related questions and 13 practice-related questions in the supplementary file. Mention if these questions are your own or if you used any valid tools or questions adopted from the relevant previous studies. Give adequate information regarding the scoring system of the variables, mentioning the highest possible aggregated score and examples of two questions (one for knowledge and one for practice).

Response: We again appreciate the reviewer. The knowledge and practice questions have been added to the supplementary file (Supplementary Table S1 and Table S2). Both knowledge and practice questions were adopted from reviewing the literature and revised according to our selection criteria. The summation scoring technique was used in computation, and their descriptive statistics, including percentiles, were observed. Then, both the knowledge and practice scores were classified according to percentile, which is evident in the existing literature (reference added). The range for the knowledge and practice scores was 1-15 and 1-11, respectively. In the main text, the section has been revised accordingly.

Statistical Analyses

The authors mentioned that they used the Mann-Whitney U test and the Kruskal-Wallis test. However, they did not mention the underlying assumptions of the tests. Moreover, the Results section also shows the χ^2 test but is not mentioned in the Methods section. Furthermore, the last line of the Results of the abstract shows the Pearson correlation coefficient, but nothing is mentioned in the Methods or Results section of the entire manuscript.

Response: We apologize for the mistake. Necessary assumptions were checked before performing statistical analysis. The Statistical Analysis section has been revised and mentions the χ^2 test and Pearson correlation coefficient. All the necessary corrections raised by the editor and reviewers have been addressed.

Results

Table 1

It is confusing as the text description of Table 1 and the title of Table 1 are different. It is recommended to use two separate tables: one for socioeconomic variables and another for the frequency distribution of the knowledge level among socioeconomic variables. Mention the knowledge- and practice-related raw scores first and then the cross-tab results. There is a major mistake in the results of Tables 1 and 2. The frequency distribution for educational status, occupation, family type, number of family members, and monthly income in Tables 1 and 2 are the same. However, the P values are different. How is this possible? Please check the results.

Response: Please accept our apology for the error that happened unconsciously. The frequency distribution for educational status, occupation, family type, number of family members, and monthly income in Tables 1 and 2 has been rechecked and revised. In addition, Table 1 has been separated into two tables (Tables 1 and 2) and presented accordingly.

Discussion

It is confusing whether the practice was for the children or how a mother takes care of their children's dental health. Mention the implications of your findings rather than just comparing the findings with previous studies. State the limitation of the study, especially the bias regarding convenient sampling. Provide a section on the public health significance of the study findings in Bangladesh.

Response: We sincerely appreciate the reviewer for these comments. The Discussion of the manuscript has been revised accordingly. The limitations have been revised in the Discussion section.

Conclusion

The Conclusion section of the study is poorly written and not focused on the findings of the study. Revise the Conclusion section to highlight your study findings.

Response: Thank you again. The Conclusion of the manuscript has been revised accordingly.

Reviewer AJ [3]

Specific Comments

There were a lot of grammatical issues and typographical errors. The manuscript needs to be edited for grammar and syntax. It is also obvious that the manuscript was not proofread adequately.

Major Comments

Abstract

- *A word is missing in the first sentence. Authors should proofread the manuscript.*
- *Keywords: Dhaka is a more appropriate keyword than Bangladesh.*
- *Under the Results in the abstract, respondents should be referred to as such and not as samples.*

Introduction

- *The global prevalence of oral diseases was stated, but authors did not capture the prevalence in the study area/country and so have not shown that oral disease is a problem. Even the global prevalence that was stated was only that of dental caries among the seven conditions that make up oral diseases as stated by the authors.*
- *The objective stated here (last sentence) comes off like the authors are assessing the knowledge and practices of oral hygiene with regard to themselves and not their children as stated in the topic.*

Methods

- *Was it permission that was given by the institutional review board or an ethical clearance?*
- *This section is quite disorganized. There is a logical flow expected in this section.*
- *Why was a nonprobability sampling technique (convenient sampling) used for this study? The sampling technique was not explained at all. This will make replicating this study difficult.*
- *I have an issue with the scoring system and the grading. Is there a reference for it? I particularly have an issue with "moderately average." It is not a standard term.*
- *The exclusion criteria are not the opposite of the inclusion criteria as stated by the authors. Exclusion criteria are those already included in the study but that are ineligible for one reason or the other.*

Results

- *In the text above Table 1, authors wrote that most respondents (39.3%) had a monthly family income of "21,000 - 40,000 taka per month." This figure (39.3%) is just over one-third of the respondents and not a majority.*
- *Table 1: What is the meaning of graduation and above? Is it graduated secondary school or graduated college?*
- *"Respectively" should be added at the end of the following sentence. "Out of 400 mothers, more than 90% knew the importance of brushing teeth while 82.3% and 80.8% of them knew the recommended frequency and appropriate time for brushing teeth."*

Discussion

- *The second sentence: the study is not evaluating parent's knowledge and practices but that of mothers.*
- *Grammatical errors and missing words*

Reference List

- *Some of the references were not cited correctly. Authors should adhere to the Vancouver referencing style.*

Round 2 Review

Reviewer BZ

The authors impressively amended the initial version of the manuscript based on the reviewers' comments. However, several issues remain unaddressed.

1. The authors should include the city in the title of the study. You can revise the title to "Knowledge and practices towards oral hygiene of children aged 5 - 9 years old: a cross-sectional study among mothers visited tertiary level hospitals in Dhaka, Bangladesh."

Response: Thanks for this suggestion. We revised the title of the manuscript accordingly as "Knowledge and practices towards oral hygiene of children aged 5 - 9 years old: a cross-sectional study among mothers visited tertiary level hospitals in Dhaka, Bangladesh."

2. Use the full form when it appears first and then use the abbreviation afterward. For example, "KP" in the abstract.

Response: Thanks again for this suggestion. We revised the title of the manuscript accordingly.

3. Please mention this statistical test in the Methods section of the abstract. You did not mention the χ^2 test and Pearson correlation.

Response: Revised the Methods section of the manuscript accordingly as "Statistical analysis including the χ^2 test and Pearson correlation test were performed. The Mann–Whitney *U* test and Kruskal–Wallis one-way ANOVA test were performed to show average knowledge and practice variations among different socio-demographics groups."

4. It is recommended to make the recommendation simple and easy to understand for the readers. Avoid duplication of the same term.

Response: Revised the Recommendation section accordingly.

5. In the sample size calculation, you used P=.58 and P=.57. Please clarify why you used those prevalences. Cite the relevant study here.

Response: The Sample Size Calculation section has been revised accordingly as "A convenient sampling technique was followed for this study. During literature search, no study was found that assessed knowledge and practice towards children's oral hygiene among Bangladeshi mothers. But, a very few studies found in other country with similar socio-demography (eg, India). Mohandas et al, 2021 in his study entitled 'Knowledge and practice of rural mothers on oral hygiene for

children' showed the prevalence of knowledge and practice were 58% and 57% respectively [4]. The sample size was calculated using the below equation.

"n =(z^2 pq)/d^2 …… (1)

"the sample size for the mother's knowledge when *P*=.58 was

"n=(〖1.96〗 ^2×0.58 × (1 - 0.58))/ 〖0.05〗 ^2=375

"Similarly, the sample size for mother's practice level when *P*=.57 was

"n=(〖1.96〗 ^2×0.57 × (1 - 0.57))/ 〖0.05〗 ^2=377

"Therefore, we initially chose a maximum of 377 as the required sample size. Considering a maximum 5% non-response rate (based on pre-testing), we rounded up this figure and selected 400 as the approximate sample size in the study."

6. Before the heading for the sociodemographic variables in the Methods section, you mention outcome measures. However, the sociodemographic variables are not your outcome variables according to your objectives. You can remove the term outcome measures from here.

Response: The heading "Outcome measure" has been removed from the revised manuscript.

7. You mentioned that you used 13 questions for the assessment of practices. Thus, according to your scoring approach, there should be a score of 1-13, but here, it is 1-11.

Response: Thank you again. We revised the error. The change is "The range for knowledge and practice score was 1 to 15, and 1 to 13 respectively."

8. Please mention the name of the software and version you used for the statistical analysis.

Response: Thank you again. We added the statistical software name with the version as "All the data management and statistical analyses were carried out through IBM SPSS Statistics 25.0."

9. Revise the sentence before Table 1. You can make it two sentences. One for family income and another for occupation.

Response: We revised the sentence accordingly as "Majority of the respondents (39.3%) had the monthly family income of 21000 - 40000 ($206.19-$392.73) Taka per month. About 13.3% mothers were involved in any paid worked activities (Table 1)."

10. There is no chi-square–related data in Table 1. Please remove the footnotes from Table 1.

Response: Removed the errors.

11. In Figure 1, it is recommended to keep the values to one decimal point for 1a and 1b.

Response: Thank you for this suggestion. We removed Figures 1c and 1d in our revised manuscript.

12. Please revise the sentence before Table 3 to give a clear meaning.

Response: We revised the sentence accordingly as "The educational status (*P*=.002) and income (*P*=.044) were significantly associated with mothers' oral hygiene practices (Table 3)."

13. You can remove the percentage symbol from the value and give it in the vertical axis title.

Response: Removed accordingly.

14. Please give the correlation results in the main manuscript or as a supplementary table.

Response: The correlation results have been given as the supplementary result. Please see Supplementary Result S6.

15. The authors overlooked the association of knowledge and practice with income and family size. Please give more details on those two points in the Discussion section.

Response: The variable family income has been addressed in the Discussion. Please see page 17 (before the Strengths and Limitation section). Family income has been discussed briefly in the Principal Findings section.

Conflicts of Interest

None declared.

References

1. Islam MH. Peer review of "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study". JMIRx Med 2025;6:e70144. [doi: 10.2196/70144]
2. Tamannur T, Das SK, Nesa A, et al. Mothers' knowledge of and practices toward oral hygiene of children aged 5-9 years in Bangladesh: cross-sectional study. JMIRx Med 2025;6:e59379. [doi: 10.2196/59379]
3. Nwankwo B. Peer review of "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study". JMIRx Med 2025;6:e70142. [doi: 10.2196/70142]
4. Mohandass B, Chaudhary H, Pal GK, Kaur S. Knowledge and practice of rural mothers on oral hygiene for children. Indian J Continuing Nurs Education 2021;22(1):39-43. [doi: 10.4103/IJCN.IJCN_7_20]

Edited by T Leung; submitted 16.12.24; this is a non–peer-reviewed article;accepted 16.12.24; published 03.02.25.

Please cite as:
Tamannur T, Das SK, Nesa A, Nahar F, Nowshin N, Binty TH, Shakil SA, Kundu SK, Siddik MAB, Rafsun SM, Habiba U, Farhana Z, Sultana H, Kamil AA, Rahman MM
Authors' Response to Peer Reviews of "Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study"
JMIRx Med 2025;6:e70145
URL: https://xmed.jmir.org/2025/1/e70145
doi:10.2196/70145

Authors' Response to Peer Reviews of "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development"

Oguzhan Serin[1], MD; Izzet Turkalp Akbasli[1], MD; Sena Bocutcu Cetin[1], MD; Busra Koseoglu[1], MD; Ahmet Fatih Deveci[2], MSc; Muhsin Zahid Ugur[2], PhD; Yasemin Ozsurekci[3], MD

[1]Department of Pediatrics, Hacettepe University Medical School, Gevher Nesibe Avenue, Altindag, Ankara, Turkey

[2]Department of Health Information Systems, University of Health Sciences, Istanbul, Turkey

[3]Department of Pediatric Infectious Diseases, Hacettepe University Medical School, Ankara, Turkey

Corresponding Author:
Izzet Turkalp Akbasli, MD
Department of Pediatrics, Hacettepe University Medical School, Gevher Nesibe Avenue, Altindag, Ankara, Turkey

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.22.24303209v1

Companion article: https://med.jmirx.org/2025/1/e71100

Companion article: https://med.jmirx.org/2025/1/e71369

Companion article: https://med.jmirx.org/2025/1/e57719

(JMIRx Med 2025;6:e71098) doi:10.2196/71098

KEYWORDS

childhood pneumonia; community-acquired pneumonia; machine learning; clinical decision support system; prognostic care decision

This is the authors' response to peer-review reports for "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development."

Round 1 Review

Anonymous [1]

General Comments

This paper [2] developed a machine learning approach that could predict community-acquired pneumonia prognosis, which is scaled into two-levels, severe or nonsevere, and identify important clinical indices, such as hypoxia, respiratory distress, age, z score of weight for age, and antibiotic usage before admission. The machine learning–based clinical decision support system tool for childhood pneumonia could provide prognostic support for case management.

Response: Thank you for your positive summary of our work. We appreciate your recognition of the machine learning tool's potential in supporting childhood pneumonia prognosis and case management.

Specific Comments

Major Comment

1. To enhance the manuscript's grounding in current research and to provide a comprehensive context for the study, the authors are recommended to incorporate an evaluation of related literature in the Introduction and Discussion sections. This could include, but not be limited to, the following studies:

- *Liu YC, Cheng HY, Chang TH, et al. Evaluation of the need for intensive care in children with pneumonia: machine learning approach. JMIR Med Inform. Jan 27, 2022;10(1):e28934. [doi: 10.2196/28934] [Medline: 35084358]*
- *Smith JC, Spann A, McCoy AB, et al. Natural language processing and machine learning to enable clinical decision support for treatment of pediatric pneumonia. AMIA Annu Symp Proc. Jan 25, 2020;2020:1130-1139. [Medline: 33936489]*
- *Kanwal K, Khalid SG, Asif M, Zafar F, Qurashi AG. Diagnosis of community-acquired pneumonia in children using photoplethysmography and machine learning-based classifier. Biomed Signal Process Control. Jan 2024;87:105367. [doi: 10.1016/j.bspc.2023.105367]*
- *Chang TH, Liu YC, Lin SR, et al. Clinical characteristics of hospitalized children with community-acquired*

pneumonia and respiratory infections: Using machine learning approaches to support pathogen prediction at admission. J Microbiol Immunol Infect. Aug 2023;56(4):772-781. [doi: 10.1016/j.jmii.2023.04.011] [Medline: 37246060]

The readers could have a more comprehensive understanding if the authors could include a concise evaluation of the prior literature in the current manuscript.

Response: Thank you for those invaluable articles. We have revised the Introduction and Discussion sections to include a concise evaluation of the recommended studies, along with other relevant literature, in order to enhance the readers' understanding and to enhance alignment with the current research landscape in this niche.

2. Considering the high stakes involved in pediatric care, particularly in intensive settings, it is critical to exam the false negative cases from the confusion matrices. Analyzing these cases for any common feature characteristics could provide insights into potential improvements in the predictive algorithm. This analysis should be clearly presented and discussed in the manuscript, emphasizing its importance in clinical decision-making.

Response: Thank you for this important suggestion. We have carefully reviewed the false negative cases and conducted an analysis to identify any common characteristics. The analysis of false negatives of the best model "Blending-2" only revealed two false negatives, underweighting clinical features comorbidities while over-relying on the absence of hypoxia. As it only included two cases, the false negatives analysis has not been included in the Results section.

3. The manuscript would benefit from a more detailed description of the cohort used in the study. Information on age, gender, and other clinical indices across the two groups (severe and nonsevere) would enable a better understanding of the study population. Additionally, providing the number of cases in each group would clarify the scope and scale of the study findings.

Response: We have added a Study Population section in the Methods, providing details on the study group and the candidate variables collected. Additionally, a Study Population Characteristics section has been included in the Results, where key variables (eg, age, respiratory distress, and leukocyte count) are compared between the nonsevere and severe level of care groups (Table 2). These updates clarify the cohort's characteristics and address your concern regarding study population details.

4. A detailed description of the data collection process is crucial for assessing the study's applicability in real-world clinical settings. The manuscript should explicitly state the following:

- *How and when clinical data, including features such as hypoxia and respiratory distress, were collected (eg, at the time of admission? or within 24 hours of admission?);*
- *The time frame considered for "antibiotic usage before admission" as relevant to the prediction model: This information is essential for replicability and for future applications of the findings in clinical workflows.*

Response: We have provided a detailed description of the variables in the revised Table 1 to enhance transparency, ensuring a better understanding of how data were collected and used for the prediction model. All clinical features were encoded by pediatricians using the unstructured initial medical records at admission. For clarity and the comprehension of readers, the phrase "...candidate features from unstructured admission notes" was added to the second paragraph under the subheading of Case Definition and Patient Selection in the Methods section. Additionally, The term "recent antibiotic usage" has been clarified to indicate oral antibiotic use prescribed before admission, specifically within the 14 days preceding hospitalization. We believe these additions provide the necessary clarity and improve the replicability of the study in real-world clinical workflows.

Reviewer E [3]

General Comments

The authors have examined the medical records for 437 patients with pneumonia and created a machine learning–based classifier to determine which patients required transfer to a tertiary care center. This subject is interesting, as the predictive power of these novel statistical techniques is high and could improve the clinical care of these patients. The authors have done thorough work describing the statistical methods used in the preprocessing of the data and model development. My primary concerns in the manuscript are the lack of clinical application description, the lack of description of the time frame of the included data elements, and the lack of description regarding the patient population and outcome of interest. The following are my point-by-point comments.

Response: Thank you for your thoughtful and detailed review of our manuscript. We appreciate your recognition of the statistical methods we used for preprocessing and model development. We acknowledge the need for improving our work in the fields that Reviewer E stated. Therefore, we have addressed each of these points as follows:

- The updated Table 1 (candidate features) provides an in-depth description of the clinical and laboratory features on how and when data collection was made (time frame), along with their clinical relevance in predicting the outcome of level of care severity. These variables were chosen based on their clinical value and ease of collection in primary care settings, allowing the model to be functional in low-resource environments.
- A new Table 2 (former Table 2 became Table 3) presents a statistical comparison between the severe and nonsevere level of care groups, focusing on the differences in demographics, clinical presentation, and laboratory values. This further highlights the factors that contribute to the outcome of interest—whether a patient requires tertiary care. The revised tables should provide a more comprehensive understanding of how the model was developed and how it applies to real-world clinical populations.

- A new subsection titled Study Population Characteristics was added under Results, where key variables were compared between groups, along with presenting the characteristics of the study population.

Specific Comments

Major Comments

Abstract

The authors use the term "case management" in the Abstract and several times in the manuscript. In this context, the authors' meaning is the decision for the escalation of care or patient transfer. However, in US-based hospital systems, case management has a different meaning, which includes largely transition to rehabilitation or nursing facilities, acquisition of home oxygen therapy, etc. I would recommend altering this term for comprehension to something like "escalation of care" or "patient triage."

Response: We acknowledge that the term "case management" may have different interpretations depending on the health care system. To avoid confusion, we will revise this term throughout the manuscript (including the main title) to either "prognostic care decision," "diagnosis and treatment," or "pneumonia management," which are more in alignment with our study's goal and contemporary research. Additionally, the Abstract has been substantially revised to align with the updated version of the manuscript.

The primary outcome of interest should be included in the Abstract.

Response: We have included a clear statement in the Abstract that the primary outcome of interest is the level of care severity, specifically focusing on the need for pediatric intensive care unit admission or advanced respiratory support.

As detailed in the Methods section, it is crucial to describe the time frame for the included variables, to know when the algorithm could be used in clinical practice.

Response: We specified the time frame for the data collection in the Abstract, in alignment with the changes made in texts and tables in the Methods section, ensuring that readers understand when the algorithm could be used in clinical practice. This will clarify the applicability of the model based on the retrospective nature of the data.

Introduction

As the goal of the algorithm in the study is to predict which patients will need transfer to tertiary care for increasing respiratory support, more of the Introduction should focus on the management of in-hospital pediatric pneumonia, challenges, and reasons for the escalation of care.

I would recommend altering the sentence that describes pneumonia as easily preventable and treatable. Several of the most complicated cases in the intensive care unit are admitted with pneumonia.

Response: Thank you for your valuable suggestions regarding the focus of the Introduction. We have revised the section to better emphasize the management of in-hospital pediatric pneumonia, including the challenges faced in recognizing and managing disease severity, as well as the reasons for escalating care. Furthermore, we have altered the sentence describing pneumonia as "easily preventable and treatable" to acknowledge the complexity of cases, particularly in intensive care settings. The revised Introduction includes the following:

1. Challenges and reasons for the escalation of care: To address this suggestion, we have expanded on the reasons for the escalation of care, providing the literature standpoint for the reasons of selecting candidate features.
2. Clarification of pneumonia's preventability and treatability: We have revised the sentence that previously described pneumonia as "easily preventable and treatable" to better reflect the complexity of the disease.
3. More focus on the management of in-hospital pediatric pneumonia: With all respect to this comment, we kindly disagree to have more focus on in-hospital pneumonia care, as it would shift the main objective of this study, which is providing prognostic care tools for primary care settings.

Methods

While great care is taken to describe the approach to data preprocessing, feature selection, and model development, I would recommend following the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis) guidelines [4], which are validated reporting recommendations for predictive models.

Response: Thank you for the insightful suggestion. We have reviewed the TRIPOD checklist and ensured that our manuscript adheres to these guidelines for transparent reporting of predictive models. We have uploaded the filled checklist under the section of "Upload Additional Material (for editors/reviewers' eyes only)."

Please provide more details regarding the hospital systems involved in this study. Are they large, academic centers or small, rural centers?

Response: Thank you for your insightful comment. In response, we have clarified the institution in the Methods section to provide better context on the hospital system involved.

For study inclusion, I am not familiar with the Integrated Management of Childhood Illness guidelines. Are these structured diagnostic codes captured in the electronic health record? Is it a computational phenotype?

Response: Thank you for raising this important point. The Integrated Management of Childhood Illness guidelines are World Health Organization recommended, providing a clinical framework for diagnosing and managing pneumonia, but they are not structured diagnostic codes in the electronic health record. Physicians manually encoded clinical features from unstructured admission notes for phenotyping, rather than using a computational phenotype. This clarification has been added to the Methods section.

Please specify what is meant by "neonatal age."

Response: We appreciate your suggestion for greater clarity. We have now specified that "neonatal age" refers to infants

younger than 28 days of life. This has been updated in the Methods section for precision.

Many of the variables included in the model are colinear. For example, age and weight are highly dependent on one another, and including both in the model can be detrimental. The feature selection methods may be able to discern this, but maybe not. I would recommend using only age and z score in the model.

Response: We appreciate your insightful comments and suggestions. It appears that including both "weight" and the "weight-for-age *z* score" derived from national reference values based on age may have caused some confusion. We have clarified this issue to ensure a more coherent presentation of the candidate features. As we only included the weight-for-age *z* score (and not weight in kilogram) in our first model, no further adjustment is required in this regard. We have retained "age" as a feature because respiratory infections and disease characteristics can vary significantly across age groups. Additionally, we kept "weight-for-age *z* score" as a separate variable, as it reflects the child's relative position among peers in the nation and serves as an indirect indicator of nutritional status.

The time frames are not stated for the variables. For example, does "hypoxia" mean hypoxia at any time during the hospitalization? On hospital admission? In the first 12 hours? This information is vital to determine the usability of the entire model. If the model uses variables available during the entire hospitalization, the predictive ability will be high, but the usability will be low. A model that can predict right when a patient is transferred to a tertiary care center that the patient will be transferred is useless. However, a model that can predict on admission, or in the first 6 - 12 hours, that a patient will require transfer is incredibly helpful. Without knowing the time frame for these variables, we cannot assess how the model could be applied in clinical practice.

Response: We thank both reviewers for raising this important point. We agree that specifying the time frames for the variables is crucial for understanding the model's applicability in clinical settings. In response, we have clarified the data collection process in the revised manuscript. All clinical features, including hypoxia and respiratory distress, are now detailed in the updated Table 1 and additional text in the Methods section under Case Definition and Patient Selection, with more emphasis on the relevant time frames of the features.

Please provide clarity regarding the study outcomes. The primary outcome is described as whether the patient was referred to a tertiary care center or not. The next sentence describes "poor prognosis" as pediatric intensive care unit admission or oxygen/ventilation support. How is this outcome used? Is this a secondary outcome? Is this describing the reason for transfer? Please clarify.

Response: Thank you for highlighting this point. We acknowledge the need to clarify the study outcomes. The primary outcome is whether the patient requires transfer to a tertiary care unit. The term "poor prognosis" refers to the reason for transfer, specifically whether the patient required pediatric intensive care unit admission or oxygen/ventilation support. This is not a separate secondary outcome, but rather the criteria used to define the primary outcome of requiring tertiary care. We have revised the manuscript to clarify that the primary outcome is the "Level of Care Severity," along with text in the Methods section to make this distinction clear.

As stated in the TRIPOD guidelines, you should present the amount of missingness in your data. It appears you used imputation methods for missing data. It is helpful to describe the amount of missing data that was imputed and the method for imputation.

Response: Thank you for your valuable comment. In accordance with the TRIPOD guidelines, we agree that reporting the amount of missing data is important for transparency. We should have mentioned our imputation method while providing details about relevant features in the first submission. We have now included a detailed description of the missing data in our revised manuscript, specifying both the percentage of missing values for each variable and the total amount of missing data. To handle missing data, we used the light gradient boosting machine algorithm as an imputation method, treating missing values as a dependent variable and predicting them based on other features to avoid bias. Individual feature weights were applied accordingly. The following features had missing values: C-reactive protein (n=34, 8.2%), albumin (n=10, 2.4%), sodium (n=8, 1.9%), aspartate aminotransferase (n=16, 3.9%), and alanine aminotransferase (n=16, 3.9%). This information has been added to the revised manuscript for clarity.

Results

There is a glaring lack of information regarding your study population. Please provide a table describing patient characteristics including demographics and the variables you used in the algorithm. Also, please provide a comparison between the patients who were transferred to a tertiary care center and those who were not.

Response: Thank you for your observation. In response, we have added a detailed description of the study population in the revised manuscript. Specifically, we have included a new subsection titled Study Population Characteristics, along with a new Table 2, which presents a comparison of the demographic and clinical characteristics between the severe and nonsevere level of care groups. We have also used appropriate statistical tests to compare the characteristics of patients requiring transfer to a tertiary care unit (severe care group) versus those who did not (nonsevere group). These additions enhance the clarity of our population description and provide a comprehensive comparison of the key variables used in our algorithm.

In imbalanced datasets, it can be more useful to measure model performance using the area under the precision-recall curve rather than the standard area under the receiver operator characteristic curve. I would recommend adding this metric.

Response: Thank you for your insightful suggestion. We agree that in the case of imbalanced datasets, the area under the precision-recall curve (PRC) can provide a more informative measure of model performance than the standard area under the receiver operating characteristic curve. In response, we have now added the PRC of all models in the performance table. We

also included a PRC plot for the blending model labeled as "Blending-2," which incorporates the top-5 highest-ranked clinical features using the optimized CatBoost, light gradient boosting machine, and extreme gradient boosting models. The new PRC plot, along with the text explaining it in the Results section, have been added to the supplementary materials to provide a more comprehensive evaluation of the model's performance on imbalanced data.

Discussion

The Discussion, overall, focuses much more on the technical details of the data curation and model development than it does on the clinical application of the model. Much of the technical details presented are also clearly explained in the Methods section and then repeated in the Discussion. I would recommend substantial revision to the Discussion section to remove redundant information that is already contained in the Methods section, as well as the addition of how this model could be applied in a clinical setting to improve the care of patients with pneumonia.

Response: We thank the reviewer for this valuable feedback. In response, we have thoroughly revised the Discussion section to reduce redundancy and place a greater focus on the clinical applications of the model, along with contemporary study inclusion. Specifically, we removed technical details that were previously repeated from the Methods section, such as the handling of imbalanced data with Synthetic Minority Oversampling Technique–Tomek, feature selection using Shapley additive explanations and recursive feature elimination with cross-validation, and detailed performance metrics for each algorithm.

In place of these technical details, we have expanded the Discussion to focus more on how the model can be used in a clinical setting to improve pneumonia care. We now highlight how the model can assist primary care physicians, especially those working in resource-limited environments, in identifying high-risk pneumonia cases that may require referral to tertiary care. We also put emphasis on predictive features (such as hypoxia, respiratory distress, age, weight *z* score, and complaint period) that are easy to assess in primary care, making the model highly practical for use in real-world clinical settings. Furthermore, we discuss the potential for the model to improve patient outcomes by facilitating timely care decisions, particularly in settings where advanced diagnostic tools may not be available.

The Discussion contains no information regarding the limitations of the study. Please describe in detail the prominent limitations of the study. These should include the use of retrospective data, including only two centers, imbalanced data, challenges with clinical implementation of the model, etc.

Response: Thank you for highlighting the need to discuss the limitations of the study in more detail. In response, we have expanded the Discussion section to include a more comprehensive account of the study's limitations. Specifically, we now address the reliance on data from a single tertiary hospital, the potential selection bias toward severe cases, the limited sample size, and the retrospective nature of the data.

The Discussion, and other areas of the manuscript, mention disease prevention several times. The goal of this study has nothing to do with the prevention of pneumonia, only the treatment of pneumonia and the prevention of associated morbidity and mortality. Please revise.

Response: Thank you for pointing out the unnecessary mentions of disease prevention in the manuscript. We agree that the primary focus of the study is on the treatment of pneumonia and the prevention of associated morbidity and mortality, not the prevention of the disease itself. We have revised the entire manuscript to eliminate any mention of disease prevention where it is not relevant and have ensured that the discussion stays focused on treatment and prognosis.

Conclusion

As it stands, the Conclusion is fairly long and does not focus only on the primary findings of the study. I would recommend trimming it to 2 - 3 sentences that focus only on the primary findings of the study, such as the feasibility of developing this type of predictive model and the potential applications of the model to clinical practice.

Response: Thank you for your feedback regarding the length and focus of the Conclusion. We agree that the Conclusion could be more concise and focused on the primary findings. Based on your suggestion, we have significantly shortened the Conclusion to focus solely on the primary findings of the study, namely, the feasibility of developing a predictive model for childhood pneumonia prognosis and its potential clinical applications. The revised Conclusion now highlights the key outcomes concisely.

Minor Comments

Methods

The authors describe that ensemble methods "significantly enhance the accuracy of classifications." Please provide a reference for this statement.

Response: We agree that providing a reference would strengthen this statement. We have now included a reference supporting our statement. Specifically, "Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: a review. Healthcare (Basel). Jun 20, 2023;11(12):1808. [doi: 10.3390/healthcare11121808] [Medline: 37372925]"

Results

Please provide numbers for those who met your primary outcome of interest (transfer to a tertiary care center).

Response: Thank you for your suggestion to provide specific numbers related to the primary outcome of interest. We have now revised the Results section to include study population characteristics along with a comparison between the severe (transferred to a tertiary care unit) and nonsevere level of care groups. The revised Results section also holds emphasis on the primary outcome of interest as follows "...Of the 437 patients analyzed, 304 patients (69.6%) met the primary outcome of being transferred required escalation of care."

Please provide a description of the time frame for patient transfer, for those who were transferred.

Response: In alignment with previous comments on the inclusion of time frames to relevant data elements, we have provided a detailed description in the updated Table 1 for candidate variables. However, our dataset does not include the timing of transfers to tertiary care units. This is recognized as a limitation of the study, and the Limitation section has been extended in this regard.

Discussion

It would be interesting to hear more regarding the use of this model in resource-limited settings and the benefits it could provide.

Response: Thank you for your valuable comments, which have already enhanced our work beyond our initial vision. We share your excitement about the future potential of this work and its possible applications.

Round 2 Review

Anonymous

I thank the authors for revising the manuscript.

Reviewer E

General Comments

The authors have conducted a single-center, retrospective study evaluating the derivation and performance of a machine learning model to predict the need for transfer to a higher level of care for childhood pneumonia. The authors were provided with a substantial amount of feedback on the original submission, and although the authors' response is detailed and comments on how all concerns were adequately addressed, the resulting manuscript is lacking in many if not most of the requested changes. The revised manuscript remains confusing to the reader and bereft of some essential elements of standard study reporting, including a basic description of the patient population and details regarding the timing of variable collection and use in the model. Due to this lack of response to the initial reviewer feedback, I am recommending rejection of this manuscript. The following are my point-by-point critiques, many of which are similar to those in my original review.

Response: We believe that these comments may stem from a review of the earlier version of our manuscript rather than the revised submission. Each specific comment raised by the reviewer was addressed in the revised manuscript, where we carefully incorporated the requested changes and clarifications. We kindly request a review of the latest version in the JMIRx system, as it reflects these substantial updates in response to the initial feedback. As the reviewer provided some additional recommendations, we made the required changes to those in our most recent manuscript. We believe there may have been a misunderstanding or an oversight, leading to the reviewer evaluating an earlier version of our manuscript. We genuinely appreciate the time and effort the reviewer has invested in helping us improve our manuscript.

Specific Comments

Abstract

First sentence: Please revise it to "Pneumonia is the leading cause of preventable mortality for children under five years of age."

Response: We have revised the first sentence of the Background section of the Abstract.

Background: The terms "case management" and "disease prevention" are still used in the Abstract. In my initial review, I recommended revising these terms to improve study clarity, and although the authors stated in their response that they replaced these terms, they remain in the Abstract. As it stands, it is not immediately clear to the reader that the purpose of the study was to provide a tool to assist bedside clinicians to determine which patients are likely to require transfer of care to a higher-level facility for pediatric pneumonia.

Response: Thank you for highlighting the importance of precise terminology in conveying the study's purpose. We have already revised the entire document to address the reviewer's initial comment/concern. We have now double-checked the revised manuscript and there is no mention of "case management" in the revised manuscript, as well as "disease prevention," that could be misunderstood by readers.

Methods: As it stands, it is confusing to the readers what was actually done in the study. It should be very apparent that the authors used a specific list of variables (please provide each in the Abstract) to predict the need for transfer to a larger institution using a specific type of machine learning model (ensemble). In the current version, this is difficult to discern.

Response: We thank your attention to the need for clarity in the Abstract. We have already addressed this concern by stating "Pediatricians encoded key clinical features from unstructured medical records based on IMCI guidelines." This line conveys that essential variables were derived from standardized guidelines without detailing each variable. Listing all variables in the Abstract would reduce clarity when considering the Abstract word limitations of this journal, especially since these variables are fully detailed in the Methods and Results sections. We believe this approach aligns with best practices for Abstract conciseness and provides sufficient information for the reader.

Results: I would be completely clear regarding the outcome your model is predicting. After reading the paper, it is understood that "pneumonia prognosis" and "severity" actually mean required transfer to a higher level of care, but it is unclear in the Abstract. I would explicitly state "predicted transfer to a higher level of care with 77% - 88% accuracy."

Response: Thank you for this valuable suggestion to improve clarity. In response, we have revised the Results section of the Abstract to explicitly state that the model predicts the need for transfer to a higher level of care, specifying the accuracy range as suggested. The revised phrasing is now "The optimized models predicted the need for transfer to a higher level of care with an accuracy of 77% - 88%..." This adjustment enhances clarity and directly conveys the model's intended outcome for readers.

Introduction

Second paragraph, fifth sentence: I would recommend revising it to "However, this preventable health problem continues to be a substantial cause of mortality, especially in underdeveloped countries and regions, due to the lack of equipment and trained human resources." There is no way to quantify it as "the most important cause of mortality."

Response: There is no mention of "the most important cause of mortality" in the revised manuscript. However, we noticed that it was in the first submission. We are deeply concerned that the reviewer's second round of comments did not provide feedback on the revised manuscript.

The term "case management" continues to be used in the Introduction, which decreases clarity for the reader.

Response: Again, these concerns have already been addressed in the revised manuscript. There is no mention of "case management." We kindly request the reviewer to read the revised version rather than the first submission that has been substantially changed after the reviewer's initial comments.

As recommended previously, I would be very specific in the Introduction that you are trying to create a tool to help bedside clinicians (typically non–intensive care physicians) decide when to transfer a patient with pneumonia to a higher level of care to prevent morbidity and mortality. As it stands, this is unclear.

Response: Thank you for this recommendation. This point was already addressed in the revised manuscript, where we clarified the study's goal in the Introduction. Please also refer to the Introduction section in the last paragraph, stating "We aimed to develop machine learning-based clinical decision support system tool for childhood pneumonia that can be used by physicians, particularly working in LMICs." However, we believe including the adjective "non-intensive care" to define these physicians in detail would improve the manuscript.

Methods

In my initial review, I asked the authors to clarify what is meant by neonatal age. In their response, they said they had revised the Methods to state specifically 28 days or fewer. However, in the first paragraph of the Methods, it continues to state "neonatal age." Please revise.

Response: Thank you for raising this point again. We did agree on this issue and corrected it in the revised manuscript as follows: "Patients younger than 28 days of age (neonatal age), older than 18 years, and those who had been hospitalized within the last 14 days were excluded." Preserving the neonatal age in this sentence is essential to emphasize that we are excluding newborn pneumonia, which requires way different clinical management and decisions.

For clarity, I would recommend restating your primary outcome to simply "required tertiary care referral." Having the outcome as severe versus nonsevere, which is defined as requiring tertiary care referral or not, adds an extra step to the thought process and can be confusing.

Response: We appreciate the recommendation to clarify the primary outcome. In the revised manuscript, we have already redefined the primary outcome to "Level of Care Severity," scaled as severe or nonsevere, and defined it as the need for referral to a tertiary care unit for intensive care or respiratory support. This phrasing preserves the conceptual framework of care severity levels while directly specifying that the outcome reflects the requirement for tertiary care referral. We believe this approach balances clarity with the study's structured outcome definitions. Additionally, this terminology is consistently used in the entire manuscript, including the Methods section, where we explicitly defined it in Table 1.

One of my largest concerns in the initial manuscript was the timing of the variables. This is crucial when determining how useful the model could be. If the elements in Table 1 are measured on admission, or in the first 6 - 12 hours of admission, the model could be very useful for patient care. If the elements were measured at any point during the hospitalization, it becomes much less useful. My worry is that the model was developed based on the elements' presence at any point, meaning if the child had fever, cough, respiratory distress, and hypoxia at hour 48, then at hour 49 the model was able to predict the patient would need transfer, and the patient was transferred at hour 50—this is not helpful to clinicians. On the other hand, if the model predicts at hour 12 that a patient needs transfer, and then at hour 50 they transfer, that is potentially very helpful to clinicians. Without these details, I cannot recommend the publication of the manuscript.

Response: Thank you for emphasizing the importance of timing in assessing the model's clinical utility again. We have already clarified this point in the revised manuscript by specifying that all variables in Table 1 were recorded at the time of admission. As stated in Table 1, these variables were extracted from initial examination documents, not from any time from the hospitalization period, reflecting the presence/measurement of variables at admission. We believe that timings are adequately mentioned by the "at admission" or "at initial examination" phrases in Table 1. Only the primary outcome "Level of Care Severity" was extracted from medical records other than the initial time point, as it is necessary to encode whether or not a patient had advanced support during their hospital stay.

It appears that the model was developed using the data from all 437 patients, and the results are presented following k-fold cross validation. It is standard practice to derive the model on a subset of the data (typically 70% - 80%) and then to test it on the remainder of the dataset to prevent overfitting and inflation of performance metrics. It does not appear that this was done. Despite having a small sample size, I believe this approach would lead to a more robust and generalizable model.

Response: Thank you for highlighting this point regarding model validation. In the revised manuscript, we confirmed that a k-fold cross-validation approach was used on the entire dataset to address the limited sample size. To mitigate concerns of overfitting and enhance model generalizability, we initially split the data, setting aside 5% as a test set to prevent data leakage. The remaining data were then used in an 85%:15% split for training and validation. This approach was chosen to maximize the utility of our sample while ensuring a robust evaluation of model performance. Please refer to the subsections named

Handling With the Imbalanced Dataset and Algorithms, where we have already addressed the reviewer's concern, in the revised manuscript from the round 1 review.

Results

The first paragraph contains many "nuts and bolts" details of model development, and these would be better positioned in the Methods section.

Response: Again, we are deeply concerned that the reviewer may not be reading the revised manuscript from the round 1 review. These concerns have already been addressed. In the revised manuscript, the Results section begins with subsection named Study Population Characteristics.

Both reviewers on the initial submission requested additional details describing the study population, and although the authors responded that they added these details, there are still none provided. It is essential to the understanding of the study results to know the characteristics of the patient population, and it should be a standard requirement for all clinical studies.

Response: We have already agreed on this issue and carefully included a substantial revision with a Study Population Characteristics subsection and a detailed Table 2, reflecting the study population adequately. Please refer to these sections, and we are prepared to address any further concerns regarding the presentation of the study population if needed.

The Shapley additive explanations value results presented in Figure 2 are valuable, but more details describing each measured factor are required. I recommend a table with each factor as rows and two columns comparing the population that did not require transfer to a tertiary care center to the population that did.

Response: Again, this concern has already been addressed by Table 2, with a basic statistical comparison between two groups including test statistics with the significance level.

An additional figure showing an area under the precision-recall curve for each model would also be interesting to the readers.

Response: On the round 1 revision, we have already included a new figure in Multimedia Appendix 2, showing the PRC. This may have been spared from the reviewer's eye.

Discussion

The Discussion spends a decent amount of space discussing the COVID-19 pandemic. While this does have some bearing on the management of childhood pneumonia, I believe the space would be better spent discussing the actual implementation of this type of algorithm. How would a primary care clinician actually use this model in practice? How would it improve upon current clinical practice? Would it be easy or difficult to incorporate into routine workflows? This would be more interesting to the readers.

Response: The revised manuscript has substantially been changed, reducing the amount of emphasis on the pandemic and carefully answering those questions that have been raised by the reviewer in the first round.

I recommend adding what the next steps of this line of research would be. How would you seek to improve the model's performance? More patient data? Additional variables?

Response: We have provided recommendations along with our limitations. Please refer to our Limitation paragraph—specifically, just before the Conclusion paragraph.

In the original submission, I recommended the authors provide a limitations section and also provided some examples. Although the authors response says they added this, there are still no limitations provided. Please provide this essential element to the Discussion.

Response: This new comment provides evidence that the reviewer was not reading the revised manuscript from the first round, because we have one relatively long paragraph dedicated to the limitations of this study. The Limitation paragraph starts with "One significant limitation of this study…" We have double-checked the JMIRx submission system, and we confidently confirm that we have uploaded the revised manuscript correctly.

Conclusion

I recommend commenting on what the next steps of this line of research would be in more specific terms.

Response: We believe that our Conclusion reflects the primary findings of the study along with its clinical importance and applicability.

Round 3 Review

Reviewer E

General Comments

The authors have conducted a single-center, retrospective study evaluating the derivation and performance of a machine learning model to predict the need for transfer to a higher level of care for childhood pneumonia. The authors were provided with a substantial amount of feedback on the original submission and have been responsive to feedback, which has resulted in a much improved manuscript. There remain several typographical and grammatical errors, which I would advise an English-grammar expert to review prior to publication, but from a scientific standpoint, I believe the manuscript is appropriate for publication.

Response: We sincerely appreciate the reviewer's recognition of the improvements made to the manuscript and their support for its scientific merit. We have carefully reviewed the manuscript for typographical and grammatical errors to ensure the highest standard of clarity and professionalism prior to publication. Thank you again for your valuable feedback that improved the quality of our work.

Specific Comments

Major Comments

1. *Details regarding the patient population have been provided in detail.*
2. *The study objectives have been clarified for readers.*
3. *The study methods are now much more reproducible.*

Response: These aspects were prioritized during the revision process, guided by the reviewers' constructive feedback, which significantly enhanced our work. Their insightful comments not only improved this manuscript but also provided valuable lessons for our future works.

References

1. Anonymous. Peer review of "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development". JMIRx Med 2025;6:e71369. [doi: 10.2196/71369]
2. Serin O, Akbasli IT, Cetin SB, et al. Predicting escalation of care for childhood pneumonia using machine learning: retrospective analysis and model development. JMIRx Med 2025;6:e57719. [doi: 10.2196/57719]
3. Rogerson C. Peer review of "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development". JMIRx Med 2025;6:e71100. [doi: 10.2196/71100]
4. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 7;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]

Abbreviations

PRC: precision-recall curve
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis

Edited by E Meinert, S Amal; submitted 09.01.25; this is a non–peer-reviewed article;accepted 09.01.25; published 04.03.25.

Please cite as:
Serin O, Akbasli IT, Cetin SB, Koseoglu B, Deveci AF, Ugur MZ, Ozsurekci Y
Authors' Response to Peer Reviews of "Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development"
JMIRx Med 2025;6:e71098
URL: https://xmed.jmir.org/2025/1/e71098
doi:10.2196/71098

Authors' Response to Peer Reviews of "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance"

Masab Mansoor[1], BS, MBA, DBA; Andrew F Ibrahim[2], BS; David Grindem[3], DO; Asad Baig[4], MD

[1]Edward Via College of Osteopathic Medicine, 4408 Bon Aire Dr, Monroe, LA, United States

[2]Texas Tech University Health Sciences Center School of Medicine, Lubbock, TX, United States

[3]Mayo Clinic, Rochester, MN, United States

[4]Department of Radiology, Columbia University Medical Center, New York, NY, United States

Corresponding Author:
Masab Mansoor, BS, MBA, DBA
Edward Via College of Osteopathic Medicine, 4408 Bon Aire Dr, Monroe, LA, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.09.24311777v1

Companion article: https://med.jmirx.org/2025/1/e73264

Companion article: https://med.jmirx.org/2025/1/e65263

(JMIRx Med 2025;6:e73258) doi:10.2196/73258

KEYWORDS

natural language processing; NLP; machine learning; ML; artificial intelligence; language model; large language model; LLM; generative pretrained transformer; GPT; pediatrics

This is the authors' response to peer-review reports for "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance."

We thank the reviewers [1] for the thoughtful and constructive feedback on our manuscript, "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance" [2]. We are grateful for the opportunity to revise and improve our work based on the insightful comments provided. Below, we provide detailed responses to the reviewers' comments and outline the changes made to the manuscript.

Comments and Responses

- *Please clarify why GPT-3.5 or GPT-4 (instead of GPT-3) was not used despite being available at the time of the study.*

Response: Thank you for highlighting this point. We have clarified that GPT-3 (DaVinci version) was selected because it was the most advanced version available during the study period. The Discussion section now also highlights the potential benefits of GPT-3.5 and GPT-4 for future studies, particularly in addressing rare or complex diagnoses.

Action taken: Added a rationale for GPT-3 selection in the Methods (Model Training and Fine-Tuning) section and expanded on the potential of GPT-3.5 and GPT-4 in the Discussion (GPT-3 vs Newer Models) section.

- *Why were racial and ethnic demographics not included? ("Data distribution gaps: No comparison of racial identity distribution between training and testing sets. Please consider adding a table or section on these demographic comparisons to ensure representation across subgroups.")*

Response: We acknowledge this limitation and have added a justification for the absence of this data. Specifically, the dataset lacked structured fields for racial or ethnic demographics due to its retrospective nature. We recommend future studies prioritize collecting this information to assess potential biases and ensure equitable performance.

Action taken: Added this explanation in the Materials and Methods (Participants and Data Collection) section.

- *Evaluation metrics: The study primarily uses specificity and sensitivity for evaluating large language model–generated responses, which may not capture the full quality of the outputs. Incorporating natural language processing metrics such as Recall-Oriented Understudy for Gisting Evaluation (ROUGE) and bilingual evaluation*

understudy (BLEU) can help assess the quality of generated responses more comprehensively. ROUGE measures the correspondence between the automatically generated response versus that of the human and what was expected. There are also issues associated with large language model generations of responses such as hallucination and the lack of attribution. Please specify or comment on how those and other issues were measured.

Response: We have included a discussion on hallucinations—where models generate inaccurate or unsupported outputs—and their implications for clinical use. Suggestions for addressing these issues, including the use of natural language processing metrics (eg, ROUGE and BLEU) and physician feedback mechanisms, have been added to the Discussion (Practical Implications) and Future Directions sections.

Action taken: Added text addressing hallucinations and quality evaluation in the relevant sections.

- *Figure 1 is mentioned but not included in the article, which affects comprehension of the study design and findings. Please include Figure 1 or provide an alternative reference to explain the content of the missing figure. Figures are helpful for readers to quickly grasp complex methodologies and findings.*

Response: Thank you for this suggestion. We have created and included a flowchart (Figure 1) summarizing the study workflow, including data collection, preprocessing, training/testing split, model fine-tuning, and evaluation steps.

Action taken: Added Figure 1 to the manuscript and referenced it in the appropriate sections.

- *Lack of clarity on potential implementation in rural health care settings: The study could be strengthened by detailing how the artificial intelligence (AI) model might be implemented in rural health care settings, including the specific challenges involved. Key considerations include the need for sufficient infrastructure (eg, electricity, internet) and the necessity of training health care providers unfamiliar with AI tools. Additionally, discussing both the potential impact (eg, improved diagnostic efficiency) and limitations (eg, handling incomplete data or overreliance on AI) would provide a more comprehensive road map for deployment in rural environments.*

Response: We have elaborated on the challenges of implementing AI tools in rural health care, including infrastructure limitations (eg, internet access, power supply) and costs. Recommendations for subsidized programs and partnerships with technology providers have been added to address these barriers.

Action taken: Expanded the Discussion (Practical Implications) section.

- *Address the lower accuracy for rare diagnoses.*

Response: We agree with this observation and have emphasized the need for targeted fine-tuning using domain-specific datasets to improve performance on rare pediatric conditions. This point is now discussed in the Discussion (Rare Diagnoses) section.

Action taken: Added text on targeted fine-tuning for rare diagnoses.

- *Normality test: The study does not address whether data normality was assessed before statistical analysis. Determining the distribution of the data is key to selecting the appropriate statistical test to analyze such data. The Kolmogorov-Smirnov test could aid in understanding data distribution, specifically testing for normality. If the data is not found to meet normality criteria, nonparametric methods should be applied. Including a data normality assessment and explaining the choice of a particular statistical test would significantly strengthen the reliability of the study.*

Response: Added data normality assessment details to Statistical Analysis section, specifying Kolmogorov-Smirnov testing and justification for parametric methods.

- *Power analysis assumptions: The assumptions underlying the power analysis are unclear, particularly regarding how specific diagnoses affect this analysis. It is advised to elaborate on the power analysis methodology, including the rationale behind sample size choices and their implications for diagnosis variability.*

Response: Expanded power analysis methodology with sample size rationale and considerations for diagnosis variability.

- *Sample size and generalizability: The sample size of 500 encounters may not adequately represent the broader pediatric population, particularly in diverse settings. Furthermore, using data from a single health care organization limits the applicability of findings to other settings. These limitations should be discussed, particularly how the validity of the results might change when it is tested with data from other health care centers. If possible, authors should mention and cite studies that reported on this effect. Additionally, future studies should consider expanding the sample size through multicenter collaborations or including data from patients with more diverse demographics to validate results across different health care environments thereby enhancing generalizability.*

Response: Enhanced discussion of sample size limitations with specific references to performance decreases across datasets (5%-15%).

- *Cross-validation across organizations: The model's reproducibility across various health care settings is not demonstrated. Evidence shows models often underperform with data from different sources. Including cross-organization validation and clearly acknowledging this limitation in the Discussion by citing relevant studies would enhance robustness. Furthermore, addressing this limitation in future work could pave the way for broader adoption and application of the model.*

Response: Added detailed Cross-Validation Limitations section citing studies showing model performance drops (12%-20%) across organizations.

- *Diagnostic exclusion or inclusion clarification: The preprocessing section does not clarify if physician diagnostics were included or excluded, leading to potential confusion for readers and impacting reproducibility. It would be helpful to know whether physician diagnostics were included in training and why. Clarifying this aspect would help standardize study replication and improve the study's transparency.*

Response: Clarified that physician-generated diagnoses were from retrospective data, not prospectively collected.

- *Data and model specifics for replicability: The study would benefit from more thorough descriptions of dataset characteristics, fine-tuning model parameters, and preprocessing methods. For validation, consider adding multicenter dataset details. Adding this information would enable other researchers to replicate and build upon the study's findings, thereby enhancing its scientific contribution.*

Response: Added comprehensive technical appendix with model specifications and implementation details.

- *Software and tools documentation: The authors describe using both Python (with scikit-learn) and IBM SPSS Statistics, but it is unclear what the software's sources are. Specifying sources for Python and scikit-learn (eg, "Python 3.8 [Python Software Foundation, Delaware, USA]") and clarifying the respective roles of Python and SPSS in the analyses would enhance transparency and allow for the reproducibility of the study.*

Response: Expanded Statistical Analysis section with rationale for test selection and metrics.

Additional Revisions

- Included a detailed Table 1 legend to clarify evaluation metrics (eg, true positive, false positive, true negative, and false negative).
- Added a sentence in the Future Directions section emphasizing the need for training programs tailored to rural health care providers.
- Corrected minor typographical errors in tables and sections for clarity.
- Expanded Introduction with relevant literature on large language models in pediatric contexts, including recent studies by Ramesh, Ghosh, and Haddad.

We hope these revisions address the reviewers' comments and improve the clarity, transparency, and quality of the manuscript. We sincerely thank the reviewers and the editorial team for their valuable feedback. Please do not hesitate to contact us with any additional comments or concerns.

References

1. Saderi D, Bender G, Olatoye T, et al. Peer review of "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance". JMIRx Med 2025;6:e73264. [doi: 10.2196/73264]
2. Mansoor M, Ibrahim AF, Grindem D, Baig A. Large language models for pediatric differential diagnoses in rural health care: multicenter retrospective cohort study comparing GPT-3 with pediatrician performance. JMIRx Med 2025;6:e65263. [doi: 10.2196/65263]

Abbreviations

AI: artificial intelligence
BLEU: bilingual evaluation understudy
ROUGE: Recall-Oriented Understudy for Gisting Evaluation

Edited by A Schwartz; submitted 28.02.25; this is a non–peer-reviewed article;accepted 28.02.25; published 19.03.25.

Please cite as:
Mansoor M, Ibrahim AF, Grindem D, Baig A
Authors' Response to Peer Reviews of "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance"
JMIRx Med 2025;6:e73258
URL: https://xmed.jmir.org/2025/1/e73258
doi:10.2196/73258

Authors' Response to Peer Reviews of "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study"

Abdul Aziz Tayoun, MPH

School of Medicine, Department of Family and Community Medicine, Jordan University, Qween Rania Street, Amman, Jordan

Corresponding Author:
Abdul Aziz Tayoun, MPH
School of Medicine, Department of Family and Community Medicine, Jordan University, Qween Rania Street, Amman, Jordan

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.03.24302286v1

Companion article: https://med.jmirx.org/2025/1/e71529

Companion article: https://med.jmirx.org/2025/1/e71531

Companion article: https://med.jmirx.org/2025/1/e57597

(JMIRx Med 2025;6:e71528) doi:10.2196/71528

KEYWORDS

periodic health examination; PHE; preventive health services; routine health checkups; Jordan; cross-sectional study

This is the author's response to peer-review reports for "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study."

Round 1 Review

Anonymous [1]

The following items were noted in this paper [2].

- *Periodic health examination (PHE) uptake: Only 27.1% of participants underwent a PHE in the last 2 years.*
- *Predictors: Significant predictors include recent visits to a primary health care facility, monthly income, and knowledge about PHEs and preventive health measures.*
- *Nonsignificant factors: Gender, marital status, smoking status, and BMI did not show a significant association with PHE uptake.*

Strengths

- *Comprehensive analysis: The study employs a robust methodology, combining descriptive, inferential, and multivariate statistical techniques to provide a thorough understanding of PHE uptake.*
- *Significant predictors identified: Key factors influencing PHE uptake were identified, offering valuable insights for health care providers and policy makers.*
- *First of its kind in Jordan: This study fills a gap in existing knowledge by being the first to investigate PHE uptake in Jordan.*

Negative Points and Areas for Improvement

Cross-Sectional Design

- *Limitation: The study's design limits the ability to establish causality.*
- *Improvement: Future research could benefit from a longitudinal approach to better establish causal relationships between the identified predictors and PHE uptake.*

Response: We acknowledge the limitation of the cross-sectional design in establishing causality and have highlighted this in the Discussion section, suggesting future longitudinal studies.

Convenience Sampling

- *Limitation: This method may introduce selection bias, and the online survey format may lead to measurement bias.*
- *Improvement: Employing a more randomized and stratified sampling method could enhance the representativeness and validity of the findings.*

Response: We have clarified the rationale for using convenience sampling due to resource constraints and have suggested more randomized methods for future studies.

Limited Generalizability

- *Limitation: Results may not be generalizable to populations outside of Jordan or those not included in the sample.*
- *Improvement: Expanding the study to include diverse populations and different geographic regions would provide a more comprehensive understanding of PHE uptake.*

Response: We understand the concern regarding generalizability. However, as the study aimed to estimate PHE uptake and its determinants specifically in Jordan, the focus on this population was intentional. For future research, we recommend conducting multinational studies, particularly in Arab countries, or performing systematic reviews or meta-analyses to obtain results that can be generalized beyond Jordan.

Survey Instrument

- *Limitation: The questionnaire's comprehensiveness and relevance to the Jordanian context might not have been fully ensured.*
- *Improvement: Pretesting the survey with a larger and more varied group, followed by adjustments based on feedback, could improve its applicability and accuracy.*

Response: We have taken steps to improve the relevance and comprehensiveness of the questionnaire by pretesting it and incorporating feedback.

Behavioral Factors

- *Limitation: The study did not find a relationship between behavioral factors and PHE uptake, which contradicts findings in other contexts.*
- *Improvement: A more detailed investigation into cultural and societal influences on health behaviors in Jordan is needed to clarify these results.*

Response: We agree that further investigation into cultural and societal influences on health behaviors in Jordan is needed and have discussed this in the manuscript.

English Language and Clarity

- *Limitation: The manuscript contains some grammatical errors and awkward phrasings, which can detract from its readability.*
- *Improvement: A thorough review and editing for language and clarity by a native English speaker or professional editor would enhance the manuscript's quality.*

Response: The manuscript has undergone a thorough review and editing process to enhance its readability and clarity.

Thank you for these excellent comments. We have thoroughly reviewed and integrated your suggestions into the main manuscript.

Reviewer AV [3]

Specific Comments

Major Comments

1. In this manuscript, write in detail about the data collection procedure.

Response: The data collection process was reviewed in detail. Please refer to the Methodology section and note that the questionnaire has been added as an appendix (see Multimedia Appendix 1).

2. Why was a convenience sampling technique employed?

Response: A convenience sampling technique was employed due to resource constraints, as the study was not funded and was conducted by a single author. This has been mentioned in the Methodology section.

3. "All collected data are treated with strict confidentiality." Some language corrections are required.

Response: We have rephrased the Ethical Consideration section to improve clarity and accuracy.

Minor Comments

There are a lot of formatting issues; many things seem copied and pasted.

Response: We have addressed the formatting issues to ensure consistency and clarity throughout the document.

Conflicts of Interest

None declared.

References

1. Anonymous. Peer review of "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study". JMIRx Med 2025;6:e71531. [doi: 10.2196/71531]
2. Tayoun AA. Determinants of periodic health examination uptake: insights from a Jordanian cross-sectional study. JMIRx Med 2025;6:e57597. [doi: 10.2196/57597]
3. Ahmed A. Peer review of "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study". JMIRx Med 2025;6:e71529. [doi: 10.2196/71529]

Abbreviations

PHE: periodic health examination

Edited by T Leung; submitted 20.01.25; this is a non–peer-reviewed article;accepted 20.01.25; published 05.02.25.

Please cite as:
Tayoun AA
Authors' Response to Peer Reviews of "Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study"
JMIRx Med 2025;6:e71528
URL: https://xmed.jmir.org/2025/1/e71528
doi:10.2196/71528

Author's Response to Peer Reviews of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis"

Hojjat Borhany, MSc

Faculty of Environmental Science, Department of Environmental Science, Informatic, and Statistics, University of Ca' Foscari Venice, Mestre (VE), Italy

Corresponding Author:
Hojjat Borhany, MSc
Faculty of Environmental Science, Department of Environmental Science, Informatic, and Statistics, University of Ca' Foscari Venice, Mestre (VE), Italy

Related Articles:

Companion article: https://www.biorxiv.org/content/10.1101/2023.06.21.545938v2

Companion article: https://med.jmirx.org/2025/1/e69895

Companion article: https://med.jmirx.org/2025/1/e69896

Companion article: https://med.jmirx.org/2025/1/e50458

(JMIRx Med 2025;6:e69894) doi:10.2196/69894

KEYWORDS

multistep fermentation; specific methane production; anaerobic digestion; kinetics study; biochar; first-order; modified Gompertz; mass balance; waste management; environment sustainability

This is the author's response to peer-review reports of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis."

Round 1 Review

Anonymous [1]

The present manuscript [2] deals with the study of the valorization of organic fractions of municipal solid waste through the production of volatile fatty acids (VFAs) and biogas. The article is interesting; in my opinion, it should be revised.

Comments

1. The presentation of the manuscript is very poor; the figures are not in the same format.

Response: The remaining figures, which included the box plots of VFA concentration, VFA/soluble chemical oxygen demand (SCOD) ratio, scheme of line, VFA and SCOD concentration, VFA weight ratio distribution, capital cost and yearly income, and biomethane content, were kept and reformulated to have the same shape. The figures outlining the kinetics study were deleted.

2. Some of the recent works should be discussed and cited in the Introduction section: [3-7].

Response: Some of the recent relevant works and studies were discussed and cited in the Introduction section as follows:

- Inyang M, Gao B, Pullammanappallil P, Ding W, Zimmerman AR. Biochar from anaerobically digested sugarcane bagasse. *Bioresour Technol*. Nov 2010;101(22):8868-8872. [doi: 10.1016/j.biortech.2010.06.088] [Medline: 20634061]
- Jung S, Shetti NP, Reddy KR, et al. Synthesis of different biofuels from livestock waste materials and their potential as sustainable feedstocks – a review. *Energy Conversion Manage*. May 15, 2021;236:114038. [doi: 10.1016/j.enconman. 2021.114038]
- Sampath P, Brijesh, Reddy KR, et al. Biohydrogen production from organic waste – a review. *Chem Eng Technol*. Jul 2020;43(7):1240-1248. [doi: 10.1002/ceat.201900400]
- Algahashm S, Qian S, Hua Y, et al. Properties of biochar from anaerobically digested food waste and its potential use in phosphorus recovery and soil amendment. *Sustainability*. Dec 10, 2018;10(12):4692. [doi:10.3390/su10124692]

3. The novelty of the work should be highlighted.

Response: We noted at the end of the Introduction and at the beginning of the Discussion that this study is novel in that it presents a strong framework for evaluating a proposal for the

financial and technical valorization of organic municipal solid waste using statistical analysis, process kinetics, mass balance, and experimental testing. Furthermore, as compared to single-step anaerobic digestion, our data showed a notably high improvement in profitability and a corresponding decrease in the payback period. In order to further close the cycle circuit and prolong the product life, we also proposed the integration of two potential future units.

4. Full stops should be removed from all subheadings.

Response: They are all removed.

5. The Results and Discussion should be written in detail with proper subheadings.

Response: The Results section was rewritten and divided into subheadings to mirror their counterparts in the Methods, and the Discussion section has the added subheadings Principal Results, Comparison With Previous Works, and Conclusion and Limitations according to the required information in the guidelines of JMIR Publications.

6. There are some typo errors; they should be rectified.

Response: They were corrected.

Reviewer GA [8]

General Comments

Generally, the manuscript should be strictly improved in English language writing and corrected for all grammatical errors throughout the whole manuscript. The author has to use a uniform style of the English language, either American or British English. Further English assistance is particularly required. Many missing articles and a lot of grammatical and punctuation errors must be corrected in the manuscript as in the corrected abstract.

Response: The abstract was prepared in an organized format and corrected for its language. We also employed English assistance. The manuscript's English was improved, and its style was harmonized with American English.

Specific Comments

This paper shows an important aspect of multiple fermentation steps for the complete utilization of municipal solid waste and conversion to useful products, which is highly recommended for circular economic sustainability worldwide. However, it needs some major revision and arrangement to allow for a better presentation of this valuable work.

Major Comments

Title

1. "Valorization of Organic Fraction of Municipal Solid Waste Through Production of Volatile Fatty Acids (VFAs) and Biogas" is a long title that should be shortened to be more concise with no abbreviations—more indicative. Suggested title: "Valorization of Organic Municipal Solid Waste for Volatile Fatty Acids and Biogas Production."

Response: It was adopted according to the guidelines for the descriptive title of the original paper: "Issue/Intervention in Demographic/Disease/Condition: Method/Study Design"; "Conversion of Organic Municipal Solid Waste to Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies with Statistical Analysis."

Abstract Section

2. General language; it must be more concise and specific.

Response: I did search for all the general language in the manuscript and tried to provide concise information on the matter.

3. Please clearly mention the take-home message and the main findings of the research.

Response: The research's primary conclusions include the development of a reliable technique for evaluating the recovery proposal for the conversion of organic solid waste into valuable products and assessing both its technical and financial viability. Furthermore, our proposal outperforms the conventional approaches in terms of economics.

4. The abstract is too long and lacks the main methodology and main experimental techniques that were carried out in this work. The author may add some hints about the main methods used before mentioning the main results.

Response: Subheadings for the background, objective, method, findings, and conclusion were added to the revised abstract. There are fewer words in the abstract overall than the 450-word limit. Additionally, some pointers regarding experimental techniques such as gas chromatography are provided, along with the kind of statistical test used to verify the significance and efficacy of the suggested process amendment. We also mentioned the use of mass flow models for the process's economic evaluation and the various kinetics models that can be used to describe biogas production.

Manuscript

5. Keywords: Words must be modified to be more informative and representative of the research interest and differ from the word in the manuscript title. Maybe add "Multi Step of Fermentation Process" or "Waste Management and Environment Sustainability."

Response: We updated the keywords to include "Multistep Fermentation," "Environment Sustainability," "Waste management," "Specific Methane Production," "Anaerobic Digestion," "Kinetics Study," "Biochar," "First-Order," "Modified Gompertz," and "Mass Balance."

6. Arrangement of the experimental work in the manuscript may be needed in the Results and Discussion accordingly.

Response: It was completed in a way that would make it easier for specialists in the field to follow the stages, and a Discussion section was included to compare the findings with earlier research, highlight the key conclusions, and clarify the research's limitations.

7. There is a lack of figures to describe the main parameter optimization steps well. Please reformulate to describe some data using figures with error bars.

Response: Our optimization procedure focused on reducing the payback period by decreasing the cost and increasing the profit from bioproducts. This was achieved through pilot tests for examining the effectiveness of the hydraulic retention time (HRT) manipulation and pretreatment in increasing the VFA yield and the integration of our process knowledge of using the fine-tuned feedstock/inoculum ratio as well as biochar addition to obtain the biogas in a cost-effective process. Detailed information and calculations regarding the mass flow analysis are available in the supplementary documents in the Excel spreadsheet named"Mass Balance.". For figures, we provided the VFA concentrations and distribution for two HRTs and a *t* test to confirm the significance of the results. Further, for biogas production, we provide results from a kinetics study showing an 8-fold increase in the hydrolysis rate and a 100% decrease in the lag phase. This brought about a small anaerobic digester working at a high organic loading rate, leading to a reasonably priced process.

8. The SD and table footnotes with the number of replicates should be noted underneath all of the given tables.

Response: For all data that was accompanied by an SD, the number of replicates was reported beneath all the given tables.

9. A mechanistic in-detail discussion is required, not just comparing your results with the previous work; justify better.

Response: The comparisons of results from similar studies were done mechanistically and in detail.

For example:

- "Because of the extra pretreatment unit in our study, our VFA yield was significantly higher than the study by valentino et al "
- "The higher hydrolysis rate was due to the destruction of the solids structure caused by bacterial enzymes and a hot alkaline solution. Additionally, we provided a higher active biomass per feedstock using a fine-tuned FS/IN ratio of 0.3 (VS basis), which was noticeably lower than the quantities (1 and 0.5) reported in similar studies "
- "due to the added fresh WS with higher digestible content and better nutrient balance than the fermented solids, the SMP value by valentino et al was higher."
- "The higher practicability than the 2 steps of bioethanol and biogas production as a result of sterilization and high bioethanol concentration requirements."
- "Our proposal is more favorable since it does not limit the VFA weight ratio distribution and does shifts the recovery route toward higher market-valued products like VFAs than single step AF + AD by Papa et al"

10. In research articles, do not include any table comparing literature results; the author can discuss the main findings in the text itself, as in Table 5.

Response: All the data in the tables comparing results were deleted, and we discussed them in the text.

11. The Conclusions section is missing in the manuscript to summarize and point out the novelty and the main findings from the research.

Response: The Conclusion was included in the manuscript and presents the main findings as follows: "To conclude, we presented a robust framework to assess a proposal for the valorization of organic waste through experimental tests, statistical analysis, and process kinetics, along with mass and energy flow analysis. The findings support considerably higher profitability and, as a result, a shorter payback period for multistep reclamation than the current single anaerobic digestion. Further, our results encourage the circular economy perspective on the conversion of OMSW into biogas and VFAs, with the pros of fewer residual solids due to reusing them in a pyrolysis line."

12. Generally speaking, in academic writing, (1) abstracts do not include abbreviations, (2) avoid articles in the title (the, a, an), and (3) avoid keywords that exist in the title.

Response: (1) Based on JMIR House Style and Guidelines, the usage of abbreviations and acronyms in the abstract section is not forbidden. Further, all author-invented abbreviations were omitted. We also stop using "AD" as an abbreviation for anaerobic digestion since it may make it ambiguous with "AD" (the reference year). In fact, keeping the number of words in the abstract within the limits is really impossible without using some of them. (2) It was avoided. (3) It was avoided to be as informative as possible.

13. As a rule of thumb, no dots in titles or subtitles as in the Experimental section, Anerobic Pilot Unities, etc.

Response: The dots were removed.

14. Multiple references should be merged, not written separately, as in "29, 30" and "23, 27"; the author may use the merge reference option in reference software.

Response: It was corrected.

15. The author may add numbers for all titles and subtitles accordingly all over the manuscript.

Response: Based on the JMIR guidelines for the author, it is not allowed to use numbering for headings and subheadings.

Minor Comments

16. The author should avoid general and well-known information, and be selective in the recent references used. May add one small paragraph to the Biological Waste Management and Environment Sustainability section.

Response: The small paragraph already discussed the current state of municipal organic waste production and treatment in the European Union. We extended it and incorporated all other information regarding environmental sustainability from some relevant sources suggested by the peer reviewer.

17. The author should clarify the main aim of the work clearly in the last paragraph of the Introduction.

Response: The main aim of this study was an assessment of multistep pretreatment acidogenic fermentation, followed by anaerobic digestion of municipal organic waste in comparison with the existing method of single anaerobic digestion in terms of financial profit and technical feasibility.

18. Do not use our, we, or us in academic writing.

Response: Based on the journal guidelines, there are no issues with using we and us in the article submitted to JMIR Publications; nevertheless, I do my best to avoid overusing these words in my manuscripts.

19. The author may mention novel applications of VFA and biogas. Mention different novel sources of biogas production.

Response: It was already mentioned in the study that biogas and VFA typically were used for energy production and biopolymer synthesis, respectively. Moreover, other sources of biogas typically were from nonbiological processes, which were beyond our scope since we focused on carbon-neutral microbiological processes.

20. The author should mention the gas chromatography type, gas injection rate, column dimensions, and the used carrier gas in the main document.

Response: It was included in the Methods section.

21. The author did not mention that flushing with nitrogen or carbon dioxide took place in anaerobic digestion while feeding reactors and how the anaerobic conditions were maintained; please mention it clearly or add the references used for the methodology.

Response: The anaerobic condition was ensured in bottles just by sealing them after filling without any flushing with nitrogen or carbon dioxide since we had known that the oxygen transfer at the surface of the waste stream was impossible as it contained high total solids and SCOD. This type of procedure was adopted in our lab and has been conducted for years.

22. Organize titles all over the manuscript.

23. Generally, the subtitles are too generic; modify them to be more indicative and precise.

Response: The subtitles were modified to be more indicative and precise.

24. "unless Saturday and Sunday" in line 208 is not important information; the suggested word "daily" is enough.

Response: It was corrected.

25. "Unite": Please correct.

Response: All units are corrected.

26. Remove the grid lines in the figures.

Response: They were removed.

27. The author has to mention the range used for the chemical oxygen demand method, and the original reference should be cited appropriately.

Response: The method for determination of soluble and solid chemical oxygen demand of the waste stream was according to the Standard Methods for Water and Wastewater. We also clearly discussed in the Methods section a proper limit of detection and reference.

28. "As can be seen": This statement is repetitive more than once in the Discussion, in lines 301, 315, and 423.

Response: Line 301 was corrected. Line 315 was corrected to be informative and avoid repetition. Line 423 was rectified in English language, and the repetitive statements were removed.

29. Figure 3 caption: Mesophilic fermentation: Please specify which stage because both of the sequential steps were called mesophilic fermentation in Figure 1.

Response: In fact, Figure 3 depicts the weight ratio distribution from the second step named mesophilic acidogenic fermentation. Surprisingly, the VFA could only be obtained from the second stage. Additionally, we modified the caption to read "VFAs weight ratio distribution for mesophilic acidogenic fermentation" and made a clear reference to Figure 1, which depicts the processes of pretreatment, acidogenic fermentation followed by mesophilic anaerobic digestion . In terms of pH and HRT, the two later procedures differ from one another substantially.

30. What is the rationale for comparing 3 days to 4.5 days for all the used systems; the author may justify why 4.5 days is better to complete with this HRT in the rest of the experiments or describe the one variable at a time optimization method that is used to determine the significant factors and the insignificant one; mention them clearly. Also, use in the Discussion the terms "significant" and "insignificant" according to the obtained P value.

Response: The values for the two HRTs to increase the VFA concentration in the outlet were selected based on our experience and process knowledge. According to this information, exceeding the HRT value of more than 3 - 5 days can bring the process into an anaerobic digestion step. As a result, the VFAs with high-added value markets are converted to biogas. Hence, the two HRTs of 3 days and 4.5 days were tried in the pilot test, knowing that the VFA concentration would either increase or decrease linearly in this local region of operation.

31. The author has to mention tables and figures in the text in their appropriate place.

Response: They were mentioned where they were referred to.

References

1. Anonymous. Peer review of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis". JMIRx Med 2025;6:e69895. [doi: 10.2196/69895]
2. Borhany H. Converting organic municipal solid waste into volatile fatty acids and biogas: experimental pilot and batch studies with statistical analysis. JMIRx Med 2025;6:e50458. [doi: 10.2196/50458]
3. Jung S, Shetti NP, Reddy KR, et al. Synthesis of different biofuels from livestock waste materials and their potential as sustainable feedstocks – a review. Energy Convers Manage 2021 May 15;236:114038. [doi: 10.1016/j.enconman.2021.114038]

4. Srivastava RK, Shetti NP, Reddy KR, Aminabhavi TM. Sustainable energy from waste organic matters via efficient microbial processes. Sci Total Environ 2020 Jun 20;722:137927. [doi: 10.1016/j.scitotenv.2020.137927] [Medline: 32208271]
5. Sampath P, Brijesh, Reddy KR, et al. Biohydrogen production from organic waste – a review. Chem Eng Technol 2020 Jul;43(7):1240-1248. [doi: 10.1002/ceat.201900400]
6. Velvizhi G, Goswami C, Shetti NP, Ahmad E, Kishore Pant K, Aminabhavi TM. Valorisation of lignocellulosic biomass to value-added products: paving the pathway towards low-carbon footprint. Fuel (Lond) 2022 Apr 1;313:122678. [doi: 10.1016/j.fuel.2021.122678]
7. Monga D, Shetti NP, Basu S, et al. Engineered biochar: a way forward to environmental remediation. Fuel (Lond) 2022 Mar 1;311:122510. [doi: 10.1016/j.fuel.2021.122510]
8. Elsalamony D. Peer review of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis". JMIRx Med 2025;6:e69896. [doi: 10.2196/69896]

Abbreviations

HRT: hydraulic retention time
SCOD: soluble chemical oxygen demand
VFA: volatile fatty acid

Edited by T Leung; submitted 10.12.24; this is a non–peer-reviewed article;accepted 10.12.24; published 04.02.25.

Please cite as:
Borhany H
Author's Response to Peer Reviews of "Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis"
JMIRx Med 2025;6:e69894
URL: https://xmed.jmir.org/2025/1/e69894
doi:10.2196/69894

Authors' Response to Peer Reviews of "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study"

Ayomide Owoyemi[1], MScPH, MD, PhD; Joanne Osuchukwu[2], MD; Megan E Salwei[3], BSc, MSc, PhD; Andrew Boyd[1], BSc, MD

[1]Department of Biomedical and Health Informatics, University of Illinois Chicago, 1919 W Taylor, Chicago, IL, United States

[2]College of Medicine, University of Cincinnati, Cincinnati, OH, United States

[3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Ayomide Owoyemi, MScPH, MD, PhD
Department of Biomedical and Health Informatics, University of Illinois Chicago, 1919 W Taylor, Chicago, IL, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69869

Companion article: https://med.jmirx.org/2025/1/e70058

Companion article: https://med.jmirx.org/2025/1/e69593

Companion article: https://med.jmirx.org/2025/1/e69594

Companion article: https://med.jmirx.org/2025/1/e69870

Companion article: https://med.jmirx.org/2025/1/e69595

Companion article: https://med.jmirx.org/2025/1/e65565

(JMIRx Med 2025;6:e69537) doi:10.2196/69537

KEYWORDS

artificial intelligence; machine learning; algorithm; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

This is the authors' response to peer-review reports for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study."

Round 1 Review

Anonymous [1]

The paper [2] presents the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF), a structured approach to guide the planning, design, development, and implementation of AI systems in health care settings. The framework is designed to address the gap between technical performance and sociotechnical factors that are essential for successful AI deployment in clinical environments.

The authors conducted a literature synthesis and a modified Delphi study involving global health care professionals to develop and refine the CASoF checklist. The checklist emphasizes the importance of considering the value proposition, data integrity, human-AI interaction, technical architecture, organizational culture, and ongoing support and monitoring, to ensure that AI tools are not only technologically sound but also practically viable and socially adaptable within clinical settings.

The study found that the successful integration of AI in health care depends on a balanced focus on both technological advancements and the sociotechnical environment of clinical settings. The CASoF represents a step forward in bridging this divide, offering a holistic approach to AI deployment that is mindful of the complexities of health care systems. The checklist aims to facilitate the development of AI tools that are effective, user-friendly, and seamlessly integrated into clinical workflows, ultimately enhancing patient care and health care outcomes.

The authors acknowledge some limitations of the study, such as the need for continuous refinement of the CASoF through iterative feedback and broader engagement with more stakeholders. Future research should aim to include an even wider array of perspectives, particularly from underrepresented regions and specialties, to enhance the framework's comprehensiveness and applicability.

Overall, the paper provides a valuable contribution to the field of AI in health care by offering a practical and comprehensive approach to the development and implementation of AI systems in clinical settings.

Reviewer AE [3]

General Comments

This paper presents the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF), a checklist intended to support the development and implementation of AI systems in health care settings. The framework is built on a comprehensive literature review and a modified Delphi study involving health care professionals globally. The manuscript addresses a significant gap in the integration of AI by emphasizing the importance of sociotechnical considerations alongside technical aspects.

Specific Comments

Major Comments

1. Clarity and structure: The manuscript could benefit from clearer explanations, particularly in the methodology section. The description of the Delphi study and literature synthesis is dense and may be difficult for readers to follow. Consider breaking down complex sentences and using more straightforward language.

Response: Thank you for this; we have addressed and improved on the clarity and description of the methodology section as requested.

2. Methodological rigor: The manuscript lacks details on the selection process for Delphi panelists and the exact methods used for data analysis. Transparency in these areas would significantly strengthen the paper. Additionally, clarify how the Delphi method was modified and the rationale behind these modifications.

Response: We have addressed the selection process and what the modification of the Delphi process involves.

3. Literature review and contextualization: The discussion section could benefit from a more critical comparison between the CASoF and existing frameworks. While the manuscript mentions other frameworks, it does not fully explore their limitations or how the CASoF overcomes these challenges. Expanding this discussion would provide a stronger justification for the CASoF's novelty and utility.

Response: We have added important comparisons with other existing frameworks/checklist and what utility the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF) has over them.

4. Checklist practicality: While the checklist is comprehensive, its length and complexity may hinder practical adoption. Consider providing a condensed version for quick reference and include examples of how the checklist can be applied in real-world scenarios.

Response: The application of the checklist in a real-world scenario has been highlighted. We appreciate the suggestion on providing a condensed version; however, we will retain the checklist in its present state and level. We created an online version to make the application easier [4].

5. Ethical considerations and bias mitigation: The manuscript discusses ethical considerations but lacks specific strategies for addressing these issues within the CASoF. Expanding this discussion would enhance the manuscript's comprehensiveness.

Response: The checklist highlights specific questions that addresses ethical considerations; this has also been better highlighted in the manuscript.

Minor Comments

6. Typographical and grammatical errors: There are minor typographical and grammatical errors throughout the manuscript that should be corrected. For instance, phrases like "revised and edited" could be simplified to "revised" for conciseness.

Response: Thanks for this comment; this has been corrected.

7. Tables and figures formatting: The tables summarizing the Delphi study results are helpful but could be more effectively formatted. Using shading or color coding to distinguish between different stages or domains would improve clarity and ease of interpretation.

Response: Thanks, this is well noted. The final formatting would be more of a decision of the publisher.

8. Recent references: Some references in the manuscript are relatively old, which is less ideal for a rapidly evolving field like AI. Where possible, the manuscript should incorporate more recent literature to support its claims and demonstrate the ongoing relevance of the topic.

Response: The references for the articles were selected based on their relevance to the topic.

Reviewer AP [5]

General Comments

This paper...is a very cohesive approach to establishing a framework for the implementation of artificial intelligence (AI).

Specific Comments

Major Comments

1. Ideally there should be information on the demographics of the expert panel.

2. Please add comments regarding equitable access for these technologies.

Response: We did not collect demographic data for the panelists except their professions.

Reviewer BH [6]

General Comments

Using artificial intelligence (AI) to add social and domain-specific steps to clinical trials is innovative. My only comment is whether the number of stages or the checklist changes if the shortlisted panelists change.

Response: This change does not affect the number of changes. The process ends when consensus is reached.

Specific Comments

Major Comments

1. I am unsure if having 38 (expert) panelists is good enough to have a robust framework.

Response: Nasa et al [7] highlighted that a panel of 30 - 50 is considered optimum for a Delphi study.

Anonymous [8]

General Comments

This paper construct a checklist to support the development and implementation of artificial intelligence (AI) in clinical settings. I only have some minor comments.

Minor Comments

1. Comparison with existing checklists: Please add a comparison with some of the existing checklists.

Response: Thank you for this; we have added the necessary comparisons.

2. Inconsistency in the number of studies: The authors initially stated that they included 20 studies, but later mentioned 23. Please clarify.

Response: This has been corrected. There were 19 studies, 3 were excluded, and then 4 were added, which gives a final total of 20.

3. Appendix visibility: The appendix is not visible.

Response: This has been corrected.

4. Abbreviation consistency: The abbreviation "IQR" appears multiple times. Please ensure it is clearly defined and used consistently.

Response: This has been corrected. Thanks.

Anonymous [9]

This paper introduces the Clinical Artificial Intelligence (AI) Sociotechnical Framework (CASoF), a checklist developed through a literature synthesis and refined by a Modified Delphi study. It aims to guide the development and implementation of AI in clinical settings, focusing on the integration of both technological performance and sociotechnical factors. The framework addresses gaps in existing frameworks by emphasizing not only technical specifications but also the broader sociotechnical dynamics essential for successful AI deployment in health care.

New approaches to reporting AI in clinical settings are crucial as AI becomes more integrated into clinical practice. However, the paper needs to address the "black box" dilemma more thoroughly. This refers to the opaque nature of AI algorithms, where the decision-making process is not easily interpretable by clinicians, leading to trust and transparency issues. Additionally, while the CASoF checklist is a valuable tool, it would benefit from a more detailed comparison to established frameworks like TRIPOD (Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis), which has been widely used in developing and validating clinical prediction models. Discussing how the CASoF complements or improves upon TRIPOD would strengthen the paper's contributions.

I suggest adding a paragraph discussing the potential roles of AI when integrated into hospital electronic health record (EHR) systems. AI could be used for the development of advanced diagnostic and prognostic tools by analyzing real-time patient data. Integration with EHRs could enhance decision-making, providing predictive analytics at the point of care and improving patient outcomes. This would help explore the broader clinical impact of AI beyond just technical integration, addressing its potential for continuous learning and optimization in health care settings.

Response: Thanks for your review, this is well noted.

References

1. Anonymous. Peer review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study". JMIRx Med 2025;6:e69869. [doi: 10.2196/69869]
2. Owoyemi A, Osuchukwu J, Salwei ME, Boyd A. Checklist approach to developing and implementing AI in clinical settings: instrument development study. JMIRx Med 2025;6:e65565. [doi: 10.2196/65565]
3. Zaki S. Peer review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study". JMIRx Med 2025;6:e70058. [doi: 10.2196/70058]
4. Owoyemi A. Clinical AI sociotechnical framework (casof). Beadaut, Inc. URL: https://bit.ly/CASOF [accessed 2025-01-23]
5. Thompson K. Peer review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study". JMIRx Med 2025;6:e69593. [doi: 10.2196/69593]
6. Saripalli S. Peer review for "Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study". JMIRx Med 2025;6:e69594. [doi: 10.2196/69594]

7. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021 Jul 20;11(4):116-129. [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]
8. Anonymous. Peer review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study". JMIRx Med 2025;6:e69870. [doi: 10.2196/69870]
9. Anonymous. Peer review for “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study". JMIRx Med 2025;6:e69595. [doi: 10.2196/69595]

Abbreviations

AI: artificial intelligence
CASoF: Clinical Artificial Intelligence Sociotechnical Framework

Edited by CN Hang, E Meinert, T Leung; submitted 02.12.24; this is a non–peer-reviewed article;accepted 02.12.24; published 20.02.25.

Please cite as:
Owoyemi A, Osuchukwu J, Salwei ME, Boyd A
Authors’ Response to Peer Reviews of “Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study”
JMIRx Med 2025;6:e69537
URL: https://xmed.jmir.org/2025/1/e69537
doi:10.2196/69537

Authors' Response to Peer Reviews of "Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study"

Sandra Bieler[1], MD; Stephan von Düring[2], MD; Damien Tagan[3], MD; Olivier Grosgurin[4], MD; Thierry Fumeaux[5], MD, MBA

[1]Médecin cheffe, Service des Urgences, Hôpital de Sion, Sion, Switzerland

[2]Faculté de Médecine de l'Université de Genève, Hôpitaux Universitaires de Genève, Genève, Switzerland

[3]Service des Soins critiques, Hôpital Riviera Chablais, Rennaz, Switzerland

[4]Service de médecine interne générale et Service des Urgences, Hôpitaux Universitaires de Genève, Genève, Switzerland

[5]Hirslanden Geneva Clinics, Geneva, Switzerland

Corresponding Author:
Sandra Bieler, MD
Médecin cheffe, Service des Urgences, Hôpital de Sion, Sion, Switzerland

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.09.28.23295699v1

Companion article: https://med.jmirx.org/2025/1/e72144

Companion article: https://med.jmirx.org/2025/1/e53276

(JMIRx Med 2025;6:e72092) doi:10.2196/72092

KEYWORDS

point-of-care ultrasonography; training program; acute respiratory failure; acute circulatory failure; emergency department

This is the authors' response to peer-review reports for "Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study."

Round 1 Review

Anonymous [1]

General Comments

This paper [2] researches an essential component of point-of care ultrasonography. As this modality is rapidly evolving, evaluation of the impact on patient management and outcomes as well as cost-effectiveness is essential. Both subjects discussed in the paper result in a highly relevant manuscript. Even though the authors discuss relevant subjects, there are some problems with the manuscript.

Specific Comments

Major Comments

1. The title of the manuscript suggests that the authors researched the impact of ultrasound after implementation. However, as stated in the Methods section, ultrasound is already used by senior physicians. Thus, the impact of ultrasound after implementation is not researched but rather the impact of ultrasound used by residents. I suggest that the authors clarify that this is a feasibility and impact study on the implementation of point-of-care ultrasound (POCUS) used by residents in the emergency department (ED) in the title and Abstract.

Response: The title has been modified according to the reviewers' indications, to highlight the fact that the study's primary aim is to validate the implementation of a training curriculum for interns in training, and not to study the effect on patient outcome.

2. The authors state that patients were not included consecutively due to logistics in phase 2. This results in a high risk of bias in the included patients. Please include in the CONSORT (Consolidated Standards of Reporting Trials)

diagram the number of patients that were eligible and were excluded based on exclusion criteria or missed.

Response: As mentioned, the patients were not fully consecutively included due to organizational reasons: an incoming patient could only be considered for inclusion if the emergency department (ED) patient flow allowed, without delaying treatment or impacting on department operations. This is mentioned in the text. However, the number of patients who could have been included is not known (no traceability of screening).

3. It is unclear how many residents were performing the ultrasound examinations included in the analysis: the Methods section state that there was only 1 resident at the ED in both phases, while in the Results section, it states that there were 12 residents trained. Please clarify.

Response: Twelve doctors were trained, but only 1 resident at a time worked in the ED during each shift, and only he or she could therefore include patients during that shift, as specified in the text. We hope that the text will clarify this point.

4. The authors state that they chose a before-and-after implementation to evaluate the effect of POCUS to avoid contamination. However, before the implementation, POCUS was already used by senior physicians, which raises the question if POCUS was indeed not used in phase 1 of the trial.

5. Interestingly, in the Discussion section, the author discussed that the publication of Msolli et al did not find an improvement of diagnostic accuracy. It would be interesting to discuss why this is the case.

Response: As suggested by the reviewer, we have added a comment on the difference in the diagnostic accuracy of point-of-care ultrasound (POCUS) in our study and in the study by Msolli et al [3].

6. In the Discussion and Conclusion, it is suggested that the use of POCUS might lead to a decrease in hospital mortality. Since this is an observational study in which, just as the authors state, a diagnostic tool rather than a therapeutic intervention is researched, this is too rash to state. Please remove this from the Conclusion and Abstract.

Response: We have modified the Conclusion to relativize the effect of implementation on mortality, which is at best indirect, as mentioned by the reviewer.

Minor Comments

Overall

7. The authors provide results with IQR; however, no ranges are given. Please describe results as mean (SD) when data are normally distributed or median (25th percentile – 75th percentile) when data are not normally distributed.

Response: As all data are not normally distributed, we have chosen to keep the IQR (25th-75th), so as not to overload the text.

8. Formatting of the full manuscript needs some attention. For example, in the Abstract, not all sentences start with a capital letter. Also, it is common in the English language to write number in full up to 20.

9. Please follow the author guidelines of the journal for reporting values and the structure of the manuscript.

Response: Formatting has been adapted according to the transmitted comments.

Title Page

10. The authors state that a clinical trial registration was done. However, it seems that they refer to a registration by a medical ethical review board. Please provide a clinical trial registration or if not applicable, remove it from the title page.

Response: We have deleted the information on registration.

Introduction

11. In the first sentence, please state the full name of "emergency department" before using the abbreviation ED.

Methods

12. Figure 1 should be formatted. The most common formatting is according to the CONSORT flow diagram.

Response: We have formatted Figure 1 according to the instructions.

Results

13. Please do not discuss the results in the Results section.

Response: We have deleted all discussions of the results in the Results section.

Discussion

14. Please end the Discussion section with the strengths and limitations. The secondary findings should be above the Strengths and Limitations section.

Response: We have moved the secondary findings to before the discussion on the strengths and limitations.

Round 2 Review

Anonymous

I would like to compliment the authors of their extensive changes to the manuscript. I have some minor comments.

Response: We thank the editor and the reviewer for their careful reading of our manuscript and for their valuable comments. We have addressed all issues raised by them and modified the text accordingly. We have uploaded a change tracking version of the manuscript, with changes highlighted in yellow.

Before-and-after design: In such a study design, the only difference between the two phases should be the implemented intervention. In IMPULSE (Impact of a Point-of-Care Ultrasound Examination), the intervention was the implementation of immediate POCUS examination by junior in-training residents managing patients in the first line, after a short structured training program. This was performed only during the postimplementation phase, and never done before. POCUS could be performed in both phases by senior experienced physicians, but later in the management of the

patient, after the initial clinical evaluation (and after the POCUS during the postimplementation phase) of the junior resident. We therefore continue to affirm that this is indeed a before-and-after study design, with a clear implementation of a changing practice. We have clarified this in all sections of the text.

We have, as suggested, included information on the residents' characteristics, as this valuable information is important for the interpretation of the study results. A new section has been added in the Methods and in the Results parts of the text.

We have put the 25th - 75th IQR range everywhere in the text and tables, as suggested.

We have removed the figure legends from the uploaded figures.

As mentioned, a change-tracking version has been uploaded as a supplementary file, with changes highlighted in yellow.

All ethics information has been grouped in a specific section in the Methods part of the text.

We have followed the guidelines on reporting results.

Minor Comments

1. I would suggest changing the sentence "However, there is still no strong evidence that the diagnostic accuracy of POCUS translates into a clinically relevant difference in patient outcomes" in the Introduction, because you also do not provide strong evidence (I do not know if we ever could provide strong evidence). I would suggest that you focus it more on the fact that the impact of using POCUS in the management of patients in the ED is still relatively unknown.

Response: We have adapted the sentence on the evidence of the clinical impact of POCUS in the Introduction, as suggested by the reviewer.

2. I would suggest to start the Discussion section with a short summary of the key findings.

Response: We have started the Discussion section with a short summary of key findings.

References

1. Anonymous. Peer review of "Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study". JMIRx Med 2025;6:e72144. [doi: 10.2196/72144]
2. Bieler S, Tagan D, Grosgurin O, Fumeaux T. Impact of a point-of-care ultrasound training program on the management of patients with acute respiratory or circulatory failure by in-training emergency department residents (IMPULSE): before-and-after implementation study. JMIRx Med 2025;6:e53276. [doi: 10.2196/53276]
3. Msolli MA, Sekma A, Marzouk MB, et al. Bedside lung ultrasonography by emergency department residents as an aid for identifying heart failure in patients with acute dyspnea after a 2-h training course. Ultrasound J 2021 Feb 9;13(1):5. [doi: 10.1186/s13089-021-00207-9] [Medline: 33559777]

Abbreviations

ED: emergency department
IMPULSE: Impact of a Point-of-Care Ultrasound Examination
POCUS: point-of-care ultrasound

Edited by E Meinert, A Schwartz; submitted 03.02.25; this is a non–peer-reviewed article;accepted 03.02.25; published 03.03.25.

Please cite as:
Bieler S, von Düring S, Tagan D, Grosgurin O, Fumeaux T
Authors' Response to Peer Reviews of "Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study"
JMIRx Med 2025;6:e72092
URL: https://xmed.jmir.org/2025/1/e72092
doi:10.2196/72092

Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study

Sandra Bieler[1], MD; Stephan von Düring[2], MD; Damien Tagan[3], MD; Olivier Grosgurin[4], MD; Thierry Fumeaux[5], MD, MBA

[1]Médecin cheffe, Service des Urgences, Hôpital de Sion, Sion, Switzerland

[2]Faculté de Médecine de l'Université de Genève, Hôpitaux Universitaires de Genève, Genève, Switzerland

[3]Service des Soins critiques, Hôpital Riviera Chablais, Rennaz, Switzerland

[4]Service de médecine interne générale et Service des Urgences, Hôpitaux Universitaires de Genève, Genève, Switzerland

[5]Hirslanden Geneva Clinics, Geneva, Switzerland

Corresponding Author:
Sandra Bieler, MD
Médecin cheffe, Service des Urgences, Hôpital de Sion, Sion, Switzerland

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.09.28.23295699v1

Companion article: https://med.jmirx.org/2025/1/e72144

Companion article: https://med.jmirx.org/2025/1/e72092

Abstract

Background: Due to its diagnostic accuracy, point-of-care ultrasound (POCUS) is becoming more frequently used in the emergency department (ED), but the feasibility of its use by in-training residents and the potential clinical impact have not been assessed.

Objective: This study aimed to assess the feasibility of implementing a structured POCUS training program for in-training ED residents, as well as the clinical impact of their use of POCUS in the management of patients in the ED.

Methods: IMPULSE (Impact of a Point-of Care Ultrasound Examination) is a before-and-after implementation study evaluating the impact of a structured POCUS training program for ED residents on the management of patients admitted with acute respiratory failure (ARF) and/or circulatory failure (ACF) in a Swiss regional hospital. The training curriculum was organized into 3 steps and consisted of a web-based training course; an 8-hour, practical, hands-on session; and 10 supervised POCUS examinations. ED residents who successfully completed the curriculum participated in the postimplementation phase of the study. Outcomes were time to ED diagnosis, rate and time to correct diagnosis in the ED, time to prescribe appropriate treatment, and in-hospital mortality. Standard statistical analyses were performed using chi-square and Mann-Whitney *U* tests as appropriate, supplemented by Bayesian analysis, with a Bayes factor (BF)>3 considered significant.

Results: A total of 69 and 54 patients were included before and after implementation of the training program, respectively. The median time to ED diagnosis was 25 (IQR 15 - 60) minutes after implementation versus 30 (IQR 10 - 66) minutes before implementation, a difference that was significant in the Bayesian analysis (BF=9.6). The rate of correct diagnosis was higher after implementation (51/54, 94% vs 36/69, 52%; *P*<.001), with a significantly shorter time to correct diagnosis after implementation (25, IQR 15 - 60 min vs 43, IQR 11 - 70 min; BF=5.0). The median time to prescribe the appropriate therapy was shorter after implementation (47, IQR 25 - 101 min vs 70, IQR 20 - 120 min; BF=2.0). Finally, there was a significant difference in hospital mortality (9/69, 13% vs 3/54, 6%; BF=15.7).

Conclusions: The IMPULSE study shows that the implementation of a short, structured POCUS training program for ED residents is not only feasible but also has a significant impact on their initial evaluation of patients with ARF and/or ACF, improving diagnostic accuracy, time to correct diagnosis, and rate of prescribing the appropriate therapy and possibly decreasing

hospital mortality. These results should be replicated in other settings to provide further evidence that implementation of a short, structured POCUS training curriculum could significantly impact ED management of patients with ARF and/or ACF.

(JMIRx Med 2025;6:e53276) doi:10.2196/53276

KEYWORDS

point-of-care ultrasonography; training program; emergency department; acute respiratory failure; acute circulatory failure

Introduction

Acute respiratory failure (ARF) and acute circulatory failure (ACF) are common causes of emergency department (ED) admissions and are associated with significant morbidity, mortality, and ED resource use. Timely and appropriate management can reduce these outcomes but depends on an efficient diagnostic workup [1]. In a high proportion of EDs around the world, patients received first-line treatment by junior in-training physicians. Traditionally, the workup is guided by history taking and physical examination, which have been shown to be inaccurate in the ED, particularly when performed by less experienced physicians [2-4]. Basic laboratory and imaging tests are often supplemented with more advanced modalities, such as transthoracic echocardiography or computed tomography (CT), at the expense of increased ED length of stay, resource use, and potential adverse events [5-7]. Point-of-care ultrasound (POCUS), performed by nonradiologists or noncardiologists, is a noninvasive bedside diagnostic tool that has been shown to be highly accurate in identifying the etiologic cause of ARF or ACF, with no significant side effects [8-20]. POCUS is now included in many training programs for emergency physicians [21-27]. However, it is still unclear if the diagnostic accuracy of POCUS translates into a clinically relevant difference in patient outcomes [18,28-33]. Despite these limitations, the American College of Physicians guidelines recommend the use of POCUS in addition to standard diagnostic procedures in patients with acute dyspnea [34,35]. In most of the published studies, POCUS was performed by trained experts who were not directly responsible for the patient and were often blinded to clinical data, which does not reflect real-life conditions where patients are initially managed by junior or in-training residents.

We designed the IMPULSE (Impact of a Point-of-Care Ultrasound Examination) study to evaluate the feasibility and impact of implementing a structured POCUS training program for in-training ED residents in the first-line management of patients admitted for ACF and/or ARF. A before-and-after implementation study design was chosen to avoid the methodological problems associated with blinding and randomization in a single-center study [35].

Methods

Study Design and Intervention

IMPULSE is a single-center, before-and-after, observational, implementation study of a structured POCUS training program for ED residents (first or second year of internal medicine training) at a regional hospital (Hôpital de Nyon, Switzerland). During the preimplementation period (phase 1), patient management was unchanged, and POCUS could only be performed on demand by trained attending physicians as part of the standard ED management implemented since 2010. Only 1 in-training ED resident per 12-hour shift participated in the study.

During the intervention phase, a group of residents in training (first and second year after graduation) were enrolled in the AURUS (Association des urgentistes et réanimateurs intéressés à l'ultrasonographie) training program, organized into 3 steps and in accordance with the European Society of Intensive Care Medicine consensus document [36-38]:

- A 20-hour, web-based course on general principles of ultrasound as well as theoretical and practical aspects of image acquisition and interpretation in transthoracic, cardiac, vascular, pulmonary, and abdominal POCUS [39]: The module includes a formal assessment of knowledge through a multiple-choice questionnaire, which must be completed to proceed to the next step.
- An 8-hour, practical, hands-on session in which POCUS examinations are performed on healthy volunteers and simulators in groups of 3 students under the supervision of an instructor, focusing on the technical aspects of obtaining interpretable images: The session includes a formal assessment of image acquisition and interpretation skills. This assessment is mandatory to proceed to the next step.
- The practice of at least 10 directly supervised POCUS full examinations, performed under real conditions in the ED: This includes a formal assessment of the ability to acquire, interpret, and integrate good-quality images into clinical management.

At the end of the training process, residents who met all training objectives were enrolled in the postimplementation phase (phase 2). Similar to phase 1, only 1 ED resident per shift participated in the study. A Sparq Ultrasound System (Philips AG Healthcare) was used for all POCUS examinations, which were performed with a 4 - 12 MHz linear probe and a 1 - 4 MHz phased array probe. POCUS was requested to be performed as soon as possible on all enrolled patients, in parallel with the clinical evaluation and according to a standardized protocol evaluating 18 specific sonographic signs (Figure 1), looking for echographic signs of pulmonary embolism, left heart failure, hypovolemic state, tamponade, pneumonia, pneumothorax, or abdominal disease. All POCUS images were recorded, and a standardized case report form was completed by the resident (Figure 2). All images were mandatorily reviewed by a POCUS-trained attending physician, directly or subsequently, to confirm the findings.

All other diagnostic procedures were used at the discretion of the clinician, including a basic POCUS performed by the

attending physician and an advanced ultrasound performed by a fully trained radiologist or cardiologist.

Figure 1. Point-of-care ultrasound (POCUS) protocol evaluating specific sonographic signs: (1) internal jugular vein; (2) to (5) anterior pulmonary view or anterior axillary line view; (6) and (7) posterobasal pulmonary view; (8) inferior vena cava; (9) parasternal short- and long-axis cardiac views; (10) apical four-chamber cardiac view; (11) subcostal cardiac view; (12) hepatorenal space; (13) splenorenal space; (14) suprapubic view; and (15) to (18) femoropopliteal veins.

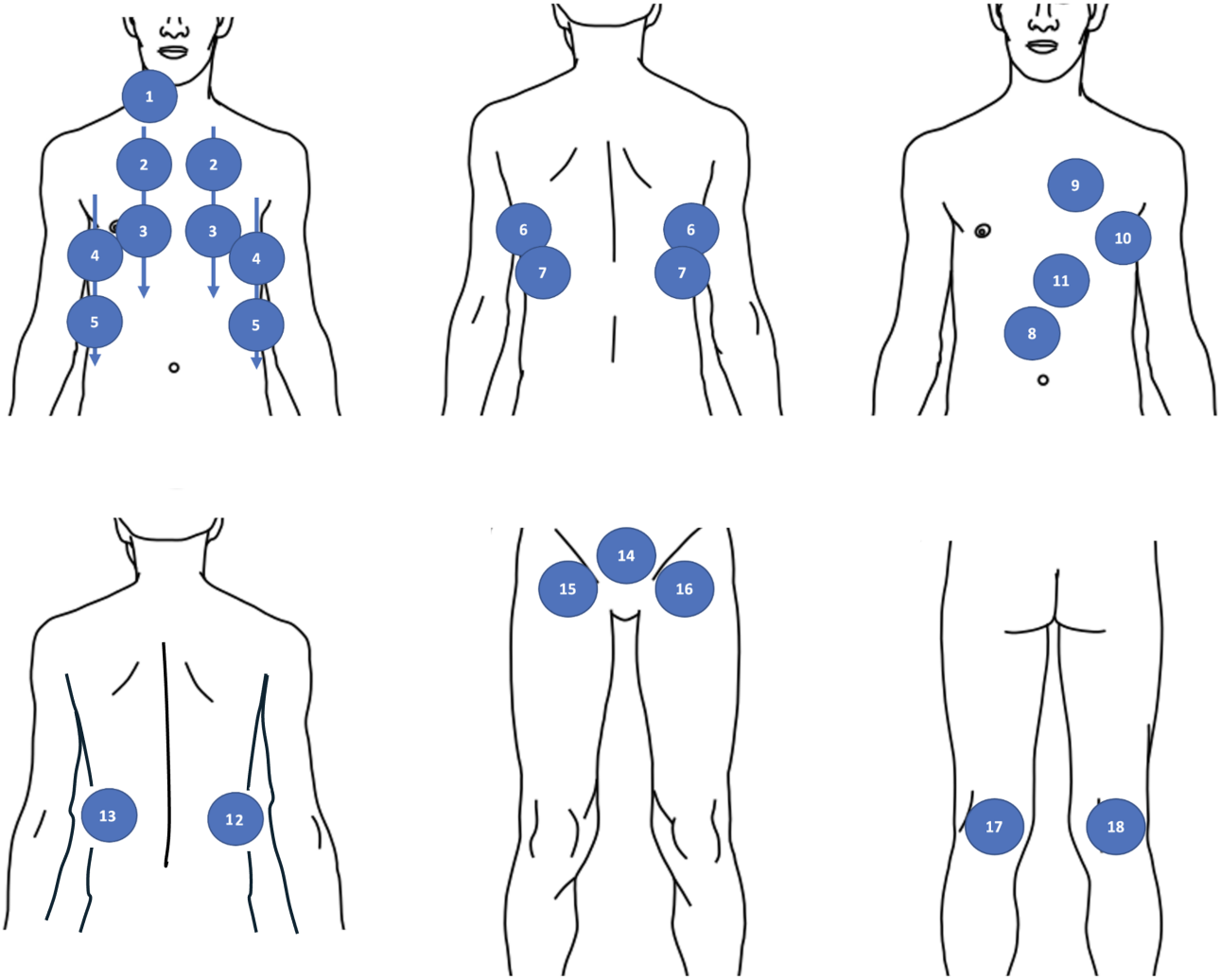


Figure 2. Case report form (adapted from the original form in French). COPD: chronic obstructive pulmonary disease; IMPULSE: Impact of a Point-of-Care Ultrasound Examination.

Case report form

Start of care : __ h__ IMPULSE identification number :

Time of diagnosis : __h__

Type of diagnosis (one or more) :

1. Pneumonia
2. Asthma/COPD exacerbation
3. Pulmonary embolism
4. Pneumothorax
5. Pericardial effusion/tamponade
6. Pleural effusion
7. Cardiac failure (acute pulmonary edema)
8. Myocardial infarction or myocarditis with cardiogenic shock
9. Septic shock
10. Gastrointestinal bleeding
11. Intraperitoneal bleeding
12. Other (specify clearly) :

Treatment prescription time : __h__

Treatment prescribed (one or more)

1. Antibiotics
2. Bronchodilators
3. Corticosteroids
4. Diuretics
5. Noninvasive ventilation (NIV)
6. Anticoagulants
7. Vasopressors
8. Coronarography
9. Abdominal surgery
10. Gastroscopy
11. Other (specify clearly ; examples : pericardial or pleural drainage, intravenous lysis, thrombectomy, arterial embolization,… .

Time of diagnosis modification (if applicable) : __h__ New diagnosis :

Comment :

Time of treatment modification (if applicable) : __h__ New treatment:

Comment :

Patient Inclusion and Exclusion Criteria

In both phases, all consecutive adult patients (aged ≥18 years) presenting with ARF and/or ACF were screened for inclusion in the study. ARF was defined by (1) the presence of either signs of respiratory distress or a respiratory rate greater than 20 breaths/min and (2) an oxygen saturation measured using pulse oximetry of <92% on room air or the need to administer oxygen to maintain a saturation of ≥92%. ACF was defined by (1) the presence of a systolic blood pressure <90 mm Hg and (2) clinical signs of hypoperfusion (agitation or altered consciousness, skin mottling, or oliguria) or hyperlactatemia (>2.0 mmol/L).

Exclusion criteria were a known or immediate diagnosis (such as ST-elevation myocardial infarction or referral for an externally determined diagnosis), the need for immediate lifesaving measures (such as cardiopulmonary resuscitation), trauma, palliative care, and patient refusal of care.

In order to preserve the organization of the ED and to favor the admission of patients for whom uninterrupted care seemed likely, the final admission of patients and the start of observation were left to the discretion of the attending physician, based on his or her assessment of the ED situation and workload.

Data Collection

On a standardized case report form, the ED resident recorded various times (start of observation, time of diagnosis, start of diagnosis-specific therapy, and end of ED stay). Diagnoses and therapies were also reported according to a specified list (Figure 1). The participating resident was equipped with an audio recorder, which was started at first contact with the patient. All recordings were kept confidential only to the investigators, who analyzed them to verify the written data reported. Based on these data, the time to diagnosis; time to prescription of targeted, appropriate treatment; and length of stay in the ED were calculated and rounded to 5-minute intervals. The hospital discharge summary was retrospectively analyzed to compare the diagnosis made during the ED stay with the final hospital diagnosis and to assess in-hospital mortality.

Statistical Analysis

All data were analyzed with the free, open-source JASP tool (University of Amsterdam). Median and IQR values are reported for descriptive statistics of continuous variables, and absolute numbers and proportions are reported for categorical variables. Differences in proportions of categorical variables between phases were analyzed by chi-square test, with a significant level set at $P<.05$. Differences in continuous variables and time intervals between phases were analyzed with a Mann-Whitney *U* test, completed by a Bayesian approach. For this analysis, the alternative hypothesis was that the time intervals would be greater in phase 1 than in phase 2, with a prior probability described by a Cauchy distribution centered around zero and with a width parameter of 1.00. This width parameter was chosen after an equivalence, Bayesian, independent-samples (2-tailed) *t* test analysis and corresponds to a probability of 50% that the effect size lies between −1.000 and 1.000. The statistical significance of the Bayesian analysis was expressed with the Bayes factor (BF), where a value between 3 and 10 is considered moderate evidence, and a value over 10 represents strong evidence. For hospital mortality comparison between the 2 phases, a Bayesian analysis was also performed, with an independent binomial analysis, with fixed rows.

Ethical Considerations

The study was approved by the regional ethics committee (Commission Cantonale d'Ethique du Canton de Vaud; protocol 194/15). Due to the observational design of the study and the fact that the practice of POCUS was already part of the usual care in the ED of the institution, a signed individual informed consent was only required for the use of the data collected for the study. Therefore, in order not to delay the management of the patients, brief verbal information was given to the patient at the beginning of the observation. Full information about the study was then given to the patient as soon as possible. Definite enrollment and data analysis were completed only after individually signed informed consent. If the patient refused to participate, then all study materials were destroyed. No compensation was provided to patients, and all data were anonymized for analysis purposes.

Results

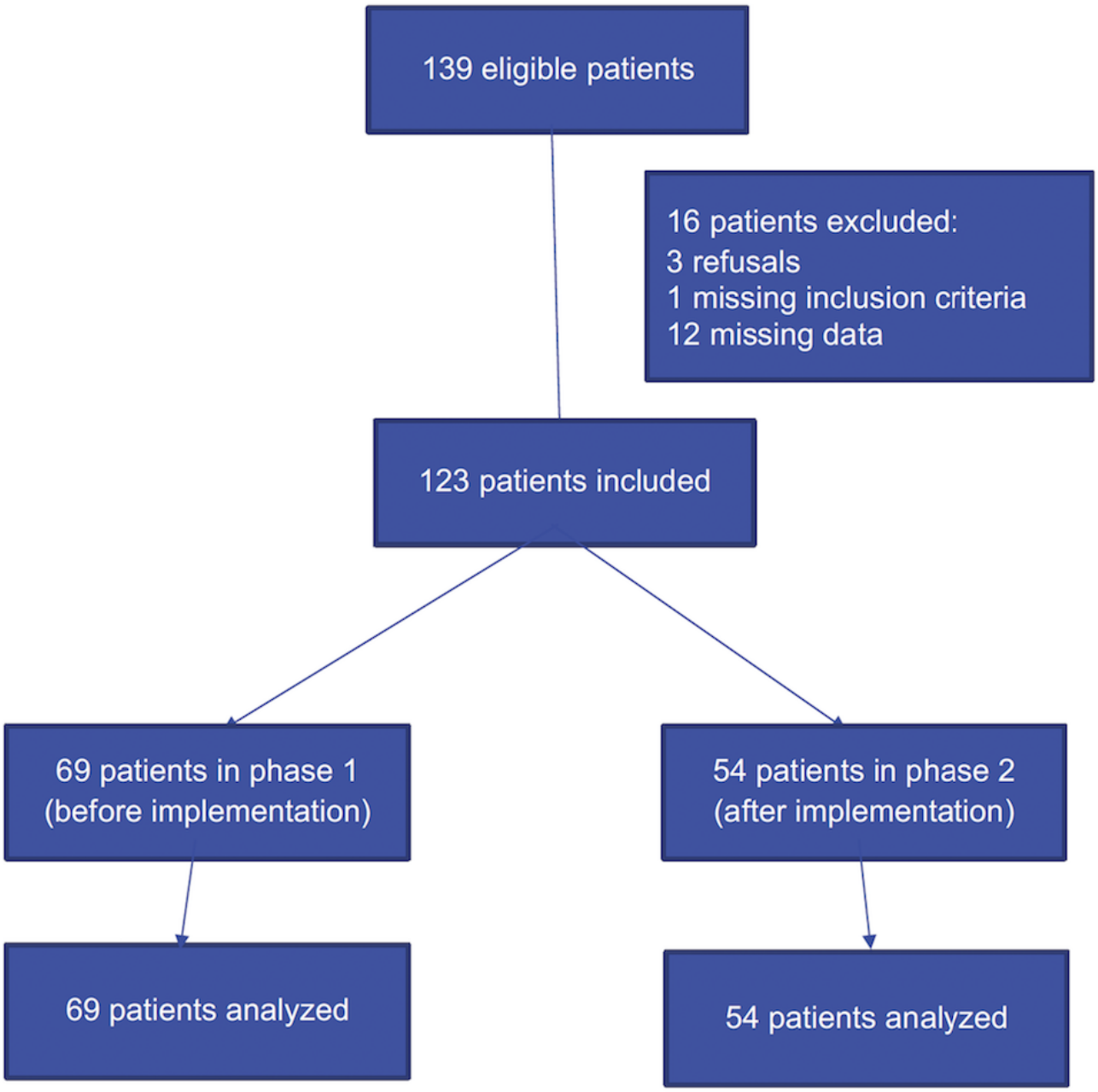
In-Training ED Residents

For ED organizational purposes, in-training residents (first or second year of training in internal medicine) were assigned to groups of 6-8 people for a 6-month rotation period. During each 12-hour shift, a resident was responsible for the first-line management of patients with ARF and/or ACF, under the supervision of an emergency medicine specialist. From September 4, 2015, to May 28, 2016 (a total of 268 days; phase 1), 14 residents participated in the observational phase, with no changes to the organization or process of usual care. Twelve interns successfully completed the AURUS training course from May 29, 2016, to September 14, 2016. Thereafter, from September 15, 2016, to February 7, 2018 (a total 511 days; phase 2), they were able to perform an immediate POCUS when managing a patient with ARF and/or ACF, which was the only difference from the observational phase 1.

Patients

During the whole study period, 139 patients were enrolled, but 3 (2.2%) patients withdrew consent to participate, 1 (0.7%) patient was excluded due to incomplete inclusion criteria, and 12 (8.6%) patients were excluded due to missing data, leaving 123 (88.5%) patients for the analysis (Figure 3). A total of 69 patients were included during phase 1 and 54 patients were included during phase 2. In the final analysis, of the 123 patients, 117 (95.1%) presented with ARF and 20 (16.3%) presented with ACF, of whom 14 (11.4%) presented with a combination of ARF (Figure 3).

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) study flowchart.



The median age of the enrolled patients was 77 (IQR 70 - 84) years, and most patients were enrolled for respiratory distress (116/123, 94.3%) and hypoxemia (117/123, 95.1%). The admission characteristics of the enrolled patients are representative of the usual patients with ARF and/or ACF who present to the ED (Table 1).

Table . Patients characteristics at admission.

	Total population (n=123)	Phase 1 (n=69)	Phase 2 (n=54)
Age (years), median (IQR)	77 (70 - 84)	78 (70 - 86)	75 (70 - 82)
Female sex, n (%)	63 (51.2)	37 (53.6)	26 (48.1)
Prehospital medicalized care, n (%)	19 (15.4)	8 (11.6)	11 (20.4)
Medical history, n (%)			
COPD[a]	35 (28.5)	21 (30.4)	14 (25.9)
Asthma	9 (7.3)	5 (7.2)	4 (7.4)
Ischemic heart disease	41 (33.3)	21 (30.4)	20 (37)
Chronic heart failure	38 (30.9)	17 (24.6)	21 (38.9)
Active or past smoking	44 (35.8)	22 (31.9)	22 (40.7)
Immunosuppressive therapy	4 (3.3)	4 (5.8)	0 (0)
Pulmonary hypertension	7 (5.7)	4 (5.8)	3 (5.6)
Chronic kidney disease	44 (35.8)	22 (31.9)	22 (40.7)
Inclusion criteria, n (%)			
Respiratory distress	116 (94.3)	64 (92.8)	52 (96.3)
Hypoxemia (SpO_2[b]<92%)	117 (95.1)	66 (95.7)	51 (94.4)
Hypotension (SBP[c]<90 mm Hg)	22 (17.9)	14 (20.3)	8 (14.8)
Clinical hypoperfusion	20 (16.3)	12 (17.4)	8 (14.8)
Admission vital signs, median (IQR)			
SpO_2 (%)	89 (83 - 92)	89 (86 - 93)	88.0 (80-92)
Respiratory rate (breaths/min)	28 (24 - 32)	28 (25 - 32)	28 (24 - 34)
Heart rate (beats/min)	100 (87 - 117)	100 (88 - 115)	105 (85 - 126)
SBP (mm Hg)	132 (112 - 152)	132 (115 - 158)	130 (110 - 152)
DBP[d] (mm Hg)	76 (61 - 89)	76 (60 - 90)	75 (63 - 89)
Laboratory values, median (IQR)			
pH	7.40 (7.35 - 7.45)	7.41 (7.35 - 7.45)	7.40 (7.36 - 7.45)
pO_2[e] (kPa)	8.2 (7.1 - 9.8)	8.3 (7.4 - 10.2	7.7 (6.7 - 9.2)
pCO_2[f] (kPa)	4.9 (4.1 - 6.3)	5.0 (4.4 - 6.0)	4.8 (3.9 - 6.8)
Lactate (mmol/L)	1.75 (1.40 - 2.75)	1.80 (1.40 - 2.85)	1.70 (1.40 - 2.28)
Creatinine (µmol/L)	104 (73 - 151)	108 (73 - 152)	98 (74 - 148)
Hemoglobin (g/L)	130 (115 - 143)	130 (114 - 144)	133 (116 - 143)
BNP[g] (ng/L)	398 (185 - 924)	267 (164 - 680)	566 (311 - 1044)
D-dimers (ug/mL)	1392 (643 - 2800)	1125 (697 - 1437)	2273 (453 - 4474)
CRP[h] (mg/L)	44 (15 - 104)	43 (15 - 95)	49 (16 - 147)

[a]COPD: chronic obstructive pulmonary disease.

[b]SpO_2: oxygen saturation.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]pO_2: partial pressure of oxygen.

[f]pCO_2: partial pressure of carbon dioxide.

[g]BNP: brain natriuretic peptide.

[h]CRP: C-reactive protein.

General ED Management

The median ED stay duration was 230 (IQR 160 - 300) minutes. During their ED stay, of the 123 patients, 98 (79.7%) had a chest x-ray, 40 (32.5%) had a chest CT scan, and 47 (38.2%) had a POCUS performed by a senior supervisor. Pneumonia was the most frequent diagnosis (n=42, 34.1%), followed by acute heart failure (n=41, 33.3%). Antibiotics (n=64, 52%) and diuretics (n=49, 39.8%) were the most frequently prescribed therapies during ED stay. Except for 2 patients (1 death and 1 home discharge), all patients were hospitalized—in half (n=58, 47.2%) of the cases, in the intensive care unit (Table 2).

Table . Emergency department (ED) management.

	Total population (n=123)	Phase 1 (n=69)	Phase 2 (n=54)
Imaging, n (%)			
Chest x-ray	98 (79.7)	65 (94.2)	33 (61.1)
Thoracic CT[a]	40 (32.5)	21 (30.4)	19 (35.2)
Abdominal CT	14 (11.4)	5 (7.2)	9 (16.7)
Abdominal ultrasound	4 (3.3)	4 (5.8)	0 (0)
Transthoracic echocardiography	3 (2.4)	2 (2.9)	1 (1.9)
POCUS[b] by senior physician	47 (38.2)	24 (34.8)	23 (42.6)
ED diagnosis, n (%)			
Pneumonia	42 (34.1)	26 (37.7)	16 (29.6)
Acute heart failure	41 (33.3)	19 (27.5)	22 (40.7)
Acute exacerbation of COPD[c]	13 (10.6)	9 (13)	4 (7.4)
Nonpulmonary sepsis	11 (8.9)	8 (11.6)	3 (5.6)
Pulmonary embolism	5 (4.1)	1 (1.4)	4 (7.4)
Pericardial effusion	3 (2.4)	0 (0)	3 (5.6)
Cardiogenic shock	2 (1.6)	1 (1.4)	1 (1.9)
Other diagnosis	6 (4.9)	5 (7.2)	1 (1.9)
Specific ED therapies, n (%)[d]			
Antibiotics	64 (52)	39 (56.5)	25 (46.3)
Diuretic therapy	49 (39.8)	24 (34.8)	25 (46.3)
Bronchodilators	27 (22)	18 (26.1)	9 (16.7)
Noninvasive ventilation	25 (20.3)	15 (21.7)	10 (18.5)
Steroids	17 (13.8)	10 (14.5)	7 (13)
Anticoagulation	14 (11.4)	5 (7.2)	9 (16.7)
Vasopressors	12 (9.8)	6 (8.7)	6 (11.1)
Patient destination after ED stay, n (%)			
Ward	58 (47.2)	36 (52.2)	22 (40.7)
ICU[e]	58 (47.2)	30 (43.5)	28 (51.9)
Other hospital (ICU or ward)	5 (4.1)	2 (2.9)	3 (5.6)
Home	1 (0.8)	1 (1.4)	0 (0)
Death in the ED	1 (0.8)	0 (0)	1 (1.9)

[a]CT: computed tomography.

[b]POCUS: point-of-care ultrasound.

[c]COPD: chronic obstructive pulmonary disease.

[d]Some patients may have received more than 1 therapy.

[e]ICU: intensive care unit.

Comparison Between Phase 1 and Phase 2

The proportion of final diagnoses retained at the end of hospitalization that confirmed the ED diagnosis was 52.2% (36/69) in phase 1 and 94.4% (51/54) in phase 2, a highly significant difference (χ^2_1=26.146, *P*<.001; Table 3).

Table . Confirmation of emergency department diagnosis during hospital diagnosis: contingency table[a].

	Diagnostic confirmed during hospital stay	
	No, n (%)	Yes, n (%)
Phase 1 (n=69)	33 (47.8)	36 (52.2)
Phase 2 (n=54)	3 (5.6)	51 (94.4)
Total (n=123)	36 (29.3)	87 (70.7)

[a]χ^2_1=26.146, *P*<.001.

Compared to phase 1, there was a statistically significant and clinically relevant decrease in the median time to final ED diagnosis in phase 2 (30, IQR 10 - 65 min vs 25, IQR 15 - 60 min; BF=9.6; Table 4).

Table . Emergency department (ED) time intervals.

	Phase 1 (n=69)	Phase 2 (n=54)	BF[a,b]	*P* value[c]
Time to final diagnosis (min), median (IQR)	30 (10 - 65)	25 (15 - 60)	9.56	.33
Time to final confirmed diagnosis (min), median (IQR)	43 (10 - 70)	25 (15 - 60)	5.02	.33
Time to administer a correct therapy (min), median (IQR)	70 (20 - 120)	47 (25 - 101)	1.96	.31
Duration of ED stay (min), median (IQR)	238 (163 - 300)	230 (160 - 275)	4.18	.42

[a]BF: Bayes factor.

[b]Alternative hypothesis: phase 1>phase 2; prior probability: Cauchy, scale 1.0.

[c]*P* value calculated with the Mann-Whitney *U* test.

When the ED diagnosis was confirmed during the hospital stay, the time to diagnosis in the ED was significantly shorter in phase 2 (25, IQR 15 - 60 min vs 43, IQR 10-70 min; BF=5.0), a difference of 18 minutes that is only moderately significant in the Bayesian analysis but clinically highly relevant. Finally, the time to order and start the most appropriate therapy was reduced from 70 (IQR 20 - 120) minutes in phase 1 to 47 (IQR 25 - 101) minutes in phase 2 (BF=2.0). There was also a reduction in the length of stay in the ED, which was significant in the Bayesian analysis, although probably not clinically relevant (Table 4).

Finally, in-hospital mortality was reduced in phase 2 (3/54, 5.6% vs 9/69, 13% in phase 1), a difference that was highly significant in Bayesian analysis (BF=16.04; Table 5).

Table . Hospital mortality: contingency table[a,b].

	Hospital mortality	
	Alive, n (%)	Dead, n (%)
Phase 1 (n=69)	60 (87)	9 (13)
Phase 2 (n=54)	51 (94.4)	3 (5.6)
Total (n=123)	111 (90.2)	12 (9.8)

[a]χ^2_1=1.93, *P*=.16.

[b]Bayesian analysis (independent multinomial analysis, with an alternate hypothesis: phase 1>phase 2): Bayes factor=16.04.

Due to the small population sample, we did not perform a formal statistical analysis of patient characteristics, components of ED management, distribution of diagnoses, and therapies administered (Tables 1 and 2). Nevertheless, we demonstrated a substantial decrease in the number of chest radiographs performed during phase 2, with an increase in the number of CT scans performed during the ED stay. In phase 1, according to the study design, a POCUS was performed by a senior attending physician in 34.8% (24/69) of the patients, whereas in phase 2, all patients had a POCUS performed by a junior attending physician, with a second POCUS performed by a senior attending physician in almost half (23/54, 42.6%) of the cases (Table 2).

Discussion

Principal Findings

The objective of the IMPULSE study was to investigate the feasibility and impact of implementing a brief, structured training program for ED residents on the management of patients admitted for ARF and/or ACF and their subsequent clinical outcomes. A before-and-after implementation design was selected to emulate the methodology of a randomized controlled trial, while mitigating the potential for contamination bias between the 2 groups. The only difference in the management of patients between the 2 phases was the immediate use of POCUS by the in-training resident in charge in the first-line treatment of the patient. The POCUS training curriculum (AURUS) was chosen for its established presence within the institution and its alignment with the updated recommendations concerning the training objectives of the current guidelines [37,38]. We hypothesized that the immediate use of POCUS by the junior physician after the short AURUS training would improve the diagnostic process, as compared by the later use by a senior physician.

The implementation of the structured, AURUS-based, POCUS program was not only associated with a significantly higher diagnostic accuracy rate but also a shorter delay of diagnosis, particularly when the ED diagnosis was later confirmed during the hospital stay. Our results also suggest that implementing a POCUS training program for in-training residents may be associated with a quicker implementation of the most appropriate therapeutic intervention, and possibly to a reduction in mortality rates, although the study design and the small sample size render the results susceptible to several potential biases. These findings align with those of a previous publication, which demonstrated that the use of POCUS by physicians of varying levels of experience was associated with an improved administration of appropriate therapies, despite no improvement in diagnostic accuracy [40]. This difference in diagnostic accuracy may be due to the more senior level of experience of the involved physicians in the published study, compared to our observation, as the diagnostic contribution of the ultrasound is probably greater for less experienced physicians.

It is also pertinent to consider some of the secondary findings of the IMPULSE study. In both phases of the study, the senior attending physician could conduct a POCUS examination; this occurred in nearly half of the cases in the postimplementation phase, a proportion that exceeds that observed in the preimplementation phase (Table 2). This may have been for verification purposes, but it is also possible that a POCUS conducted by a junior physician may prompt more experienced physicians to perform it with greater frequency, as a ripple effect. Similarly, although this finding should be interpreted with caution, there was a reduction in the number of chest x-rays performed during phase 2 (61.1% of patients only). This suggests that the POCUS may be used in place of this examination. Conversely, the number of CT scans performed during phase 2 was higher, which could be interpreted in two ways. It could be a negative effect of the POCUS, whereby supervisors performed more CT scans to confirm or reject a diagnosis made by their junior colleagues. The observed increase in the number of POCUS examinations performed by supervisors suggests that this may be a more positive effect. POCUS provides a more comprehensive assessment of the clinical situation, leading to a more appropriate use of advanced diagnostic modalities. Subsequent studies will likely address these findings and may confirm these trends, while providing clarification regarding the causes of the observed increase in CT scan use.

Our results show that the reported intervention is not only feasible but also that it has an impact on the clinical management process and possibly on the patient outcome. To the best of our knowledge, these data represent the inaugural demonstration of the clinical impact of a POCUS training program for ED residents. If replicated, they could substantiate the implementation of POCUS in conjunction with history taking and clinical examination by ED residents as a primary diagnostic tool.

Strengths and Limitations

The IMPULSE study has several notable strengths. The study design reflects the typical circumstances observed in most EDs, wherein patients are initially managed by junior physicians under the guidance of more experienced, senior medical professionals. The characteristics of the included patients and the diagnoses made in the ED demonstrate that this study sample is representative of the population of interest for the use of POCUS, with significant associated morbidity and mortality. The before-and-after study design circumvents the contamination bias observed in several previously published studies. The initial phase reflects the typical practice of most EDs, wherein POCUS is conducted by senior physicians at a relatively late stage, serving as a control for the subsequent postimplementation phase. Interestingly, the rate of inaccurate ED diagnosis during the phase 1 reflects the usual diagnostic accuracy for the management of patients who present to the ED [41-43].

The signal of a clinically relevant impact on the patient outcome is an interesting finding, as morbidity and mortality are the usual end points of choice for ED interventional studies. As POCUS is not a therapeutic procedure, the effect on outcome can only be driven by a quicker and more appropriate administration of efficient therapies. Therefore, our findings of quicker and more accurate diagnosis may explain the reduction of hospital mortality that was evidenced in our small population sample.

It is important to consider the limitations of the IMPULSE study, including the lack of randomization. However, as there is a risk of contamination between the two arms of a randomized controlled trial, we therefore elected to use a before-and-after implementation design as the optimal method to achieve quasi-randomization of patients to limit this risk. A cluster randomization of multiple centers with successive implementation would likely have been the optimal design in this situation; however, it was not feasible to organize. A second limitation is the single-center design and the limited sample of included patients, despite a lengthy recruitment period, particularly in phase 2, with 1 included patient every 9 days. This illustrates the challenges inherent in conducting

single-center studies in smaller institutions lacking dedicated clinical research resources. Notwithstanding this significant limitation, the studied population is representative of the typical patients with ARF and/or ACF admitted to the majority of EDs globally, as evidenced by their characteristics and corresponding diagnoses. It would be prudent to reproduce our results in other clinical settings, with the inclusion of a larger sample of patients, before any firm conclusion can be made regarding the impact of implementing a POCUS training program for in-training ED residents. These limitations do not affect the fundamental conclusions of the presented results.

Conclusion

In conclusion, the IMPULSE study demonstrates that a brief, structured training program for ED residents is both feasible and enables them to use POCUS as a primary tool for the initial management of patients presenting with ARF and/or ACF. The deployment of POCUS by these less experienced physicians may be associated with an increase in diagnostic accuracy, comparable to that observed in published data on POCUS use by experienced ED physicians. Furthermore, it may be associated with a reduction in the time required for in-training residents to reach a correct diagnosis and with a more rapid and appropriate prescription of a specific therapy, which may result in a decrease in hospital mortality. The results of the IMPULSE study also validate the AURUS training curriculum, demonstrating that this structured, stepwise approach to training is not only feasible but also efficient. These results must be replicated and validated in other settings with larger patient samples. However, the methodology presented herein is appropriate for limiting the issues of blinding and randomization in the study of such diagnostic tools and may be used by future studies.

Conflicts of Interest

None declared.

References

1. Ray P, Birolleau S, Lefort Y, et al. Acute respiratory failure in the elderly: etiology, emergency diagnosis and prognosis. Crit Care 2006;10(3):R82. [doi: 10.1186/cc4926] [Medline: 16723034]
2. Leuppi JD, Dieterle T, Koch G, et al. Diagnostic value of lung auscultation in an emergency room setting. Swiss Med Wkly 2005 Sep 3;135(35-36):520-524. [doi: 10.4414/smw.2005.10886] [Medline: 16323069]
3. Wipf JE, Lipsky BA, Hirschmann JV, et al. Diagnosing pneumonia by physical examination: relevant or relic? Arch Intern Med 1999 May 24;159(10):1082-1087. [doi: 10.1001/archinte.159.10.1082] [Medline: 10335685]
4. Mulrow CD, Lucey CR, Farnett LE. Discriminating causes of dyspnea through clinical examination. J Gen Intern Med 1993 Jul;8(7):383-392. [doi: 10.1007/BF02600079] [Medline: 8410400]
5. Collins SP, Lindsell CJ, Storrow AB, Abraham WT, ADHERE Scientific Advisory Committee, Investigators and Study Group. Prevalence of negative chest radiography results in the emergency department patient with decompensated heart failure. Ann Emerg Med 2006 Jan;47(1):13-18. [doi: 10.1016/j.annemergmed.2005.04.003] [Medline: 16387212]
6. Al Aseri Z. Accuracy of chest radiograph interpretation by emergency physicians. Emerg Radiol 2009 Mar;16(2):111-114. [doi: 10.1007/s10140-008-0763-9] [Medline: 18779982]
7. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. N Engl J Med 2007 Nov 29;357(22):2277-2284. [doi: 10.1056/NEJMra072149] [Medline: 18046031]
8. Lichtenstein DA, Mezière GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. Chest 2008 Jul;134(1):117-125. [doi: 10.1378/chest.07-2800] [Medline: 18403664]
9. Lichtenstein D. FALLS-protocol: lung ultrasound in hemodynamic assessment of shock. Heart Lung Vessel 2013;5(3):142-147. [Medline: 24364005]
10. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. Anesthesiology 2004 Jan;100(1):9-15. [doi: 10.1097/00000542-200401000-00006] [Medline: 14695718]
11. Copetti R, Soldati G, Copetti P. Chest sonography: a useful tool to differentiate acute cardiogenic pulmonary edema from acute respiratory distress syndrome. Cardiovasc Ultrasound 2008 Apr 29;6:16. [doi: 10.1186/1476-7120-6-16] [Medline: 18442425]
12. Agricola E, Bove T, Oppizzi M, et al. "Ultrasound comet-tail images": a marker of pulmonary edema: a comparative study with wedge pressure and extravascular lung water. Chest 2005 May;127(5):1690-1695. [doi: 10.1378/chest.127.5.1690] [Medline: 15888847]
13. Chavez MA, Shams N, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. Respir Res 2014 Apr 23;15(1):50. [doi: 10.1186/1465-9921-15-50] [Medline: 24758612]
14. Cibinel GA, Casoli G, Elia F, et al. Diagnostic accuracy and reproducibility of pleural and lung ultrasound in discriminating cardiogenic causes of acute dyspnea in the emergency department. Intern Emerg Med 2012 Feb;7(1):65-70. [doi: 10.1007/s11739-011-0709-1] [Medline: 22033792]

15. Volpicelli G, Lamorte A, Tullio M, et al. Point-of-care multiorgan ultrasonography for the evaluation of undifferentiated hypotension in the emergency department. Intensive Care Med 2013 Jul;39(7):1290-1298. [doi: 10.1007/s00134-013-2919-7] [Medline: 23584471]
16. Zanobetti M, Poggioni C, Pini R. Can chest ultrasonography replace standard chest radiography for evaluation of acute dyspnea in the ED? Chest 2011 May;139(5):1140-1147. [doi: 10.1378/chest.10-0435] [Medline: 20947649]
17. Xirouchaki N, Magkanas E, Vaporidi K, et al. Lung ultrasound in critically ill patients: comparison with bedside chest radiography. Intensive Care Med 2011 Sep;37(9):1488-1493. [doi: 10.1007/s00134-011-2317-y] [Medline: 21809107]
18. Laursen CB, Sloth E, Lambrechtsen J, et al. Focused sonography of the heart, lungs, and deep veins identifies missed life-threatening conditions in admitted patients with acute respiratory symptoms. Chest 2013 Dec;144(6):1868-1875. [doi: 10.1378/chest.13-0882] [Medline: 23948720]
19. Sasmaz MI, Gungor F, Guven R, Akyol KC, Kozaci N, Kesapli M. Effect of focused bedside ultrasonography in hypotensive patients on the clinical decision of emergency physicians. Emerg Med Int 2017;2017:6248687. [doi: 10.1155/2017/6248687] [Medline: 28357139]
20. Gartlehner G, Wagner G, Affengruber L, et al. Point-of-care ultrasonography in patients with acute dyspnea: an evidence report for a clinical practice guideline by the American College of Physicians. Ann Intern Med 2021 Jul;174(7):967-976. [doi: 10.7326/M20-5504] [Medline: 33900798]
21. Abbasi S, Farsi D, Hafezimoghadam P, Fathi M, Zare MA. Accuracy of emergency physician-performed ultrasound in detecting traumatic pneumothorax after a 2-h training course. Eur J Emerg Med 2013 Jun;20(3):173-177. [doi: 10.1097/MEJ.0b013e328356f754] [Medline: 22828649]
22. Bustam A, Noor Azhar M, Singh Veriah R, Arumugam K, Loch A. Performance of emergency physicians in point-of-care echocardiography following limited training. Emerg Med J 2014 May;31(5):369-373. [doi: 10.1136/emermed-2012-201789] [Medline: 23428721]
23. Jones AE, Tayal VS, Kline JA. Focused training of emergency medicine residents in goal-directed echocardiography: a prospective study. Acad Emerg Med 2003 Oct;10(10):1054-1058. [doi: 10.1111/j.1553-2712.2003.tb00574.x] [Medline: 14525737]
24. Moore CL, Rose GA, Tayal VS, Sullivan DM, Arrowood JA, Kline JA. Determination of left ventricular function by emergency physician echocardiography of hypotensive patients. Acad Emerg Med 2002 Mar;9(3):186-193. [doi: 10.1111/j.1553-2712.2002.tb00242.x] [Medline: 11874773]
25. Mandavia DP, Aragona J, Chan L, Chan D, Henderson SO. Ultrasound training for emergency physicians--a prospective study. Acad Emerg Med 2000 Sep;7(9):1008-1014. [doi: 10.1111/j.1553-2712.2000.tb02092.x] [Medline: 11043996]
26. Filopei J, Siedenburg H, Rattner P, Fukaya E, Kory P. Impact of pocket ultrasound use by internal medicine housestaff in the diagnosis of dyspnea. J Hosp Med 2014 Sep;9(9):594-597. [doi: 10.1002/jhm.2219] [Medline: 24891227]
27. Counselman FL, Sanders A, Slovis CM, Danzl D, Binder LS, Perina DG. The status of bedside ultrasonography training in emergency medicine residency programs. Acad Emerg Med 2003 Jan;10(1):37-42. [doi: 10.1111/j.1553-2712.2003.tb01974.x] [Medline: 12511313]
28. Bellone A, Etteri M, Maino C, Bonetti C, Natalizi A. The role of bedside ultrasound in the diagnosis and outcome of patients with acute respiratory failure. Emerg Care J 2013;9(1):e2. [doi: 10.4081/ecj.2013.e2]
29. Kanji HD, McCallum J, Sirounis D, MacRedmond R, Moss R, Boyd JH. Limited echocardiography-guided therapy in subacute shock is associated with change in management and improved outcomes. J Crit Care 2014 Oct;29(5):700-705. [doi: 10.1016/j.jcrc.2014.04.008] [Medline: 24857642]
30. Pirozzi C, Numis FG, Pagano A, Melillo P, Copetti R, Schiraldi F. Immediate versus delayed integrated point-of-care-ultrasonography to manage acute dyspnea in the emergency department. Crit Ultrasound J 2014;6(1):5. [doi: 10.1186/2036-7902-6-5] [Medline: 24940478]
31. Jones AE, Tayal VS, Sullivan DM, Kline JA. Randomized, controlled trial of immediate versus delayed goal-directed ultrasound to identify the cause of nontraumatic hypotension in emergency department patients. Crit Care Med 2004 Aug;32(8):1703-1708. [doi: 10.1097/01.ccm.0000133017.34137.82] [Medline: 15286547]
32. Atkinson PRT, McAuley DJ, Kendall RJ, et al. Abdominal and Cardiac Evaluation with Sonography in Shock (ACES): an approach by emergency physicians for the use of ultrasound in patients with undifferentiated hypotension. Emerg Med J 2009 Feb;26(2):87-91. [doi: 10.1136/emj.2007.056242] [Medline: 19164614]
33. Zieleskiewicz L, Lopez A, Hraiech S, et al. Bedside POCUS during ward emergencies is associated with improved diagnosis and outcome: an observational, prospective, controlled study. Crit Care 2021 Jan 22;25(1):34. [doi: 10.1186/s13054-021-03466-z] [Medline: 33482873]
34. Cid X, Canty D, Royse A, et al. Impact of point-of-care ultrasound on the hospital length of stay for internal medicine inpatients with cardiopulmonary diagnosis at admission: study protocol of a randomized controlled trial-the IMFCU-1 (Internal Medicine Focused Clinical Ultrasound) study. Trials 2020 Jan 8;21(1):53. [doi: 10.1186/s13063-019-4003-2] [Medline: 31915052]
35. Prager R, Wu K, Bachar R, et al. Blinding practices during acute point-of-care ultrasound research: the BLIND-US meta-research study. BMJ Evid Based Med 2021 Jun;26(3):110-111. [doi: 10.1136/bmjebm-2020-111577] [Medline: 33177166]

36. Tagan D, Fumeaux T, Beaulieu Y, Association des urgentistes et réanimateurs intéressés par l'ultrasonographie. Innovative concept in ultrasonography training targeted for the intensivist using e-learning and simulation [Article in French]. Rev Med Suisse 2015 Apr 1;11(468):785-786. [doi: 10.53738/REVMED.2015.11.468.0785] [Medline: 26021141]
37. Azarnoush K, Guechi Y, Schmutz T, Peyrony O, Fumeaux T, Ribordy V. Point-of-care ultrasonography, update on practices and a concept of implementation in an emergency department [Article in French]. Rev Med Suisse 2019 May 8;15(650):984-989. [doi: 10.53738/REVMED.2019.15.650.0984] [Medline: 31066531]
38. Expert Round Table on Ultrasound in ICU. International expert statement on training standards for critical care ultrasonography. Intensive Care Med 2011 Jul;37(7):1077-1083. [doi: 10.1007/s00134-011-2246-9] [Medline: 21614639]
39. Online courses: basic (whole body). POCUS Academy. URL: https://pocus.academy/en_GB/lesson-types/online/basic-level [accessed 2025-02-14]
40. Msolli MA, Sekma A, Marzouk MB, et al. Bedside lung ultrasonography by emergency department residents as an aid for identifying heart failure in patients with acute dyspnea after a 2-h training course. Ultrasound J 2021 Feb 9;13(1):5. [doi: 10.1186/s13089-021-00207-9] [Medline: 33559777]
41. Peng A, Rohacek M, Ackermann S, et al. The proportion of correct diagnoses is low in emergency patients with nonspecific complaints presenting to the emergency department. Swiss Med Wkly 2015 May;145:w14121. [doi: 10.4414/smw.2015.14121] [Medline: 25741894]
42. Sikka R, Tommaso LH, Kaucky C, Kulstad EB. Diagnosis of pneumonia in the ED has poor accuracy despite diagnostic uncertainty. Am J Emerg Med 2012 Jul;30(6):881-885. [doi: 10.1016/j.ajem.2011.06.006] [Medline: 21855251]
43. Johnson T, McNutt R, Odwazny R, Patel D, Baker S. Discrepancy between admission and discharge diagnoses as a predictor of hospital length of stay. J Hosp Med 2009 Apr;4(4):234-239. [doi: 10.1002/jhm.453] [Medline: 19388065]

Abbreviations:

ACF : acute circulatory failure
ARF : acute respiratory failure
AURUS: Association des urgentistes et réanimateurs intéressés à l'ultrasonographie
BF : Bayes factor
CT: computed tomography
ED: emergency department
IMPULSE: Impact of a Point-of-Care Ultrasound Examination
POCUS: point-of-care ultrasound

Edited by A Schwartz, E Meinert; submitted 02.10.23; peer-reviewed by Anonymous; revised version received 06.01.25; accepted 30.01.25; published 03.03.25.

Please cite as:
Bieler S, von Düring S, Tagan D, Grosgurin O, Fumeaux T
Impact of a Point-of-Care Ultrasound Training Program on the Management of Patients With Acute Respiratory or Circulatory Failure by In-Training Emergency Department Residents (IMPULSE): Before-and-After Implementation Study
JMIRx Med 2025;6:e53276
URL: https://xmed.jmir.org/2025/1/e53276
doi:10.2196/53276

Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study

Tahazid Tamannur[1], MPH; Sadhan Kumar Das[1], MPH; Arifatun Nesa[2], MPH; Foijun Nahar[1], MPH; Nadia Nowshin[1], MPH; Tasnim Haque Binty[1], MPH; Shafiul Azam Shakil[2], MPH; Shuvojit Kumar Kundu[3], MPH; Md Abu Bakkar Siddik[4], MPH; Shafkat Mahmud Rafsun[5], MPH; Umme Habiba[6], MPH; Zaki Farhana[7], MS; Hafiza Sultana[1], MPhil; Anton Abdulbasah Kamil[8], PhD; Mohammad Meshbahur Rahman[9], MS

[1]Department of Health Education, National Institute of Preventive and Social Medicine, Dhaka, Bangladesh

[2]Department of Public Health and Hospital Administration, National Institute of Preventive and Social Medicine, Mohakhali, Dhaka, Bangladesh

[3]Directorate General of Health Services, Ministry of Health & Family Welfare, Government of the People's Republic of Bangladesh, Dhaka, Bangladesh

[4]School of the Environment, Nanjing University, Nanjing, China

[5]Dental Speciality Center, Dhaka, Bangladesh

[6]BRAC James P Grant School of Public Health, BRAC University, Dhaka, Bangladesh

[7]Credit Information Bureau, Bangladesh Bank, Dhaka, Bangladesh

[8]Department of Business Administration, Istanbul Gelisim University, Istanbul, Turkey

[9]Department of Biostatistics, National Institute of Preventive and Social Medicine, Dhaka, Bangladesh

Corresponding Author:
Mohammad Meshbahur Rahman, MS
Department of Biostatistics, National Institute of Preventive and Social Medicine, Dhaka, Bangladesh

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.04.05.24305403v1

Companion article: https://med.jmirx.org/2025/1/e70142

Companion article: https://med.jmirx.org/2025/1/e70144

Companion article: https://med.jmirx.org/2025/1/e70145

Abstract

Background: Healthy oral hygiene is crucial for overall health and well-being. Parents' dental care knowledge and practices affect their children's oral health.

Objective: This study examined mothers' knowledge and practices regarding their children's oral hygiene through a cross-sectional survey.

Methods: This cross-sectional survey was conducted from January 1 to December 31, 2022, in Dhaka, Bangladesh. Mothers' knowledge and practices regarding their children's oral hygiene were assessed through a semistructured questionnaire. Statistical analyses, including the χ^2 test and Pearson correlation test, were performed. The Mann-Whitney *U* and Kruskal-Wallis 1-way ANOVA tests were also used to show the average variations in knowledge and practices among different sociodemographic groups.

Results: Of 400 participants, the mean age of mothers was 30.94 (SD 5.15) years, and 388 (97%) were of the Muslim faith, 347 (86.8%) were housewives, and 272 (68%) came from nuclear families. A total of 165 (41.3%) participants showed good knowledge of their children's oral hygiene, followed by 86 (21.5%) showing moderately average knowledge, 75 (18.8%) showing average knowledge, and 74 (18.5%) showing poor knowledge. A total of 182 (45.5%) mothers had children with good oral hygiene practices, followed by mothers with children who had average (n=78, 19.5%), moderately average (n=75, 18.8%), and poor (n=65, 16.3%) oral hygiene practices. The mother's knowledge level was significantly associated with age (*P*=.01), education (*P*<.001), family size (*P*=.03), and monthly income (*P*<.001). On the other hand, educational status (*P*=.002) and income (*P*=.04) were significantly associated with the mother's practices regarding their children's oral hygiene. Nonparametric analysis revealed that mothers who were older (mean knowledge score: 12.13, 95% CI 10.73-13.54 vs 11.21, 95% CI 10.85-11.58; *P*=.01), with a

bachelor's degree or higher (mean knowledge score: 12.93, 95% CI 12.55 - 13.31 vs 9.66, 95% CI 8.95 - 10.37; *P*<.001), who were working mothers (mean knowledge score: 12.30, 95% CI 11.72 - 12.89 vs 11.45, 95% CI 11.17 - 11.73; *P*=.03), and who had a higher family income (mean knowledge score: 12.49, 95% CI 12.0 - 12.98 vs 10.92, 95% CI 10.48 - 11.36; *P*<.001) demonstrated significantly higher levels of oral health knowledge. Conversely, good oral hygiene practices were significantly associated with higher maternal education (mean practice score: 6.88, 95% CI 6.54 - 7.22 vs 6.01, 95% CI 5.63 - 6.40; *P*<.001) and family income (mean practice score: 6.77, 95% CI 6.40 - 7.14 vs 5.96, 95% CI 5.68 - 6.24; *P*=.002). The mother's knowledge was also significantly and positively correlated (Pearson correlation coefficient *r*=0.301; *P*<.001) with their children's oral hygiene practices, shown by both the Pearson chi-square (χ^2=25.2; *P*<.001) test and correlation coefficient.

Conclusions: The mothers' knowledge and their children's oral hygiene practices were inadequate. The mother's age, education level, family size, and monthly income significantly influenced their knowledge level. Children's oral hygiene habits were significantly associated with family income and the mother's educational status. This underscores the need for educational programs, accessible dental care services, oral health education in the curriculum, media and technology involvement in oral health educational campaigns, and proper research and monitoring.

(JMIRx Med 2025;6:e59379) doi:10.2196/59379

KEYWORDS

mothers' knowledge and practices; oral hygiene; child oral health; Bangladesh

Introduction

According to the World Health Organization, dental caries, periodontal disease, tooth loss, mouth cancer, oro-dental trauma, noma, and congenital defects including cleft lip and palate are classified as oral diseases. Oral health issues are prevalent in low-income nations owing to poor socio–educational-economic circumstances [1]. In terms of general health and well-being, there is a significant connection between oral health and overall health [2,3]. It impacts individuals' capacity to do tasks, communicate, and engage in social interactions. Thus, it has an impact on both the physical and psychological aspects of an individual [4]. Most common oral health problems and conditions can be readily avoided by establishing suitable oral hygiene routines, such as twice daily brushing with the best toothbrush, using fluoride-containing toothpaste, and using the proper brushing technique [5]. Other preventive measures include eating a balanced diet low in free sugar, going to the dentist regularly for exams, and receiving treatment for illnesses when they are still in the early stages [6]. It can be minimized by practicing good oral hygiene habits, such as brushing and flossing teeth and visiting the dentist frequently [7].

Worldwide, over 2 billion individuals have dental caries in their permanent teeth, while 514 million children have dental caries in their primary teeth [8]. Early childhood caries (ECC) in children have been linked mostly to poor dental hygiene. Infants and toddlers with significant plaque accumulation were more likely to experience severe ECC and caries from birth to toddlerhood [9]. ECC has several causes, including excessive sugar intake, poor dental hygiene, inadequate fluoride exposure, and enamel abnormalities [10]. So, the development of caries and the acquisition of infection are substantially influenced by diet and feeding habits.

The children in Bangladesh have various infections and disorders [11-13]. Poor oral health is another prevalent health problem among them, which is still neglected [13]. As parents are the major caregivers, their involvement is crucial in the maintenance and development of excellent oral health in children, such as teaching healthy eating and drinking habits [14]. In addition, several factors impact the dental health of children, including the mother's level of education, the mother's work situation, and her understanding of oral hygiene [15]. The adoption of good oral health practices in children is influenced by the parents', and particularly the mother's, oral health knowledge, attitudes, and awareness [16]. An Indian study found that the oral hygiene quality of children aged 12 years was shown to be significantly influenced by their mother's oral hygiene knowledge [17]. Children with high rates of dental caries and low rates of fillings were found to have parents with inadequate oral health literacy, according to another study [18]. As a result, it is essential for parents, and particularly mothers, to have awareness about oral health. Scholars argued that a mother's knowledge about oral health and the consequence of adequate dental hygiene has a beneficial impact on their children's dental well-being and adherence to dental care practices [19,20].

Research on dental caries awareness among parents in Pakistan has found low levels of knowledge about oral hygiene standards [21]. A study conducted in India on the oral health status of children aged 3 - 6 years and their mother's oral health–related knowledge, attitude, and practices found most mothers had a medium level of knowledge, an average level of attitude, and a high level of practices regarding oral health [22]. Another study in Malaysia on parental knowledge and practices in preschool children's oral health found that the majority had good knowledge [23]. Numerous studies have been conducted globally regarding parents' or mothers' oral hygiene knowledge and practices, but these have been insufficient, particularly among mothers in Bangladesh. There is a lack of research investigating the extent to which mothers are aware of and follow oral hygiene practices. Hence, this study aimed to assess mothers' level of oral hygiene knowledge and practices regarding their 5- to 9-year-old children.

Methods

This study followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines to prepare the manuscript, and the STROBE checklist is provided in Multimedia Appendix 1.

Ethical Considerations

Permission to conduct this study was given by the institutional review board of the National Institute of Preventive and Social Medicine (NIPSOM), Bangladesh (Ref NIPSOM/IRB/2017/09). The Shaheed Suhrawardy Medical College Hospital and Dhaka Dental College Hospital provided the necessary documentation. Both written and verbal consent were taken before initiating the interview. Participants received an overview of the study's goals, and those who consented were eventually included. No compensation was given to the participants, and data anonymity was strictly maintained.

Study Setting and Participants

This cross-sectional study was conducted from January 1 to December 31, 2022, in two tertiary-level hospitals in Dhaka South named Shaheed Suhrawardy Medical College Hospital and Dhaka Dental College Hospital in Dhaka City. Mothers of children aged 5-9 years who visited these tertiary hospitals were interviewed through a semistructured questionnaire.

Study Pretesting

To observe the overall scenario including questionnaire information, possible sampling techniques, and approximate nonresponse rate in the study, we first performed a pretest of the study. The pretesting was conducted among 50 mothers of children aged 5 - 9 years in the Sapporo Dental College & Hospital located at Dhaka North.

Sampling Technique and Sample Size

A convenience sampling technique was followed for this study. During the literature search, no study was found that assessed the knowledge and practices toward children's oral hygiene among Bangladeshi mothers. However, a study was found from India with a similar sociodemography. Mohandass et al [20] showed that the prevalence of adequate knowledge and practices were 58% and 57%, respectively. The sample size for this study was calculated using the below equation.

$$(1) n=z2pqd2$$

The sample size when P=.58 for the mother's knowledge was:

$$n=1.962\times0.58\times(1-0.58)0.052=375$$

Similarly, the sample size when P=.57 for the mother's practice level was:

$$n=1.962\times0.57\times(1-0.57)0.052=377$$

Therefore, we initially chose a maximum of 377 as the required sample size. Considering a maximum 5% nonresponse rate (based on pretesting), we rounded up this figure and selected 400 as an approximate sample size for the study.

Selection Criteria

The inclusion criteria for this study were mothers of Bangladeshi nationality who were living in Dhaka for at least 1 year, mothers of children aged 5 - 9 years, and mothers who provided consent and agreed to participate in the study. The exclusion criteria for the study were mothers who were not Bangladeshi but currently living in Dhaka, mothers of children older than 10 years or younger than 5 years, and mothers younger than 21 years or older than 48 years.

Sociodemographic Variables

Respondents' sociodemographic variables such as age (21-48 years), religion (Muslim, non-Muslim), educational status (up to primary, secondary, higher secondary, and bachelor's degree or higher), occupational status (housewife, working), family type (nuclear, joint), family size (<5 persons, ≥5 persons), and monthly family income (≤20,000 BDT, 20,001 - 40,000 BDT, ≥40,001 BDT; a currency exchange rate of 101.85 BDT=US $1 was used) were the independent variables in this study.

Measurement of Knowledge and Practice

The study used 15 variables to assess mothers' knowledge and 13 to assess their children's practices related to oral hygiene (Multimedia Appendices 2 and 3). Both knowledge and practice questions were adopted from the existing literature and revised according to our selection criteria. The summation scoring technique was used in computation, and the descriptive statistics, including percentiles, were observed. The range for the knowledge and practice scores were 1-15 and 1-13, respectively. According to the percentile approach, knowledge was classified into four levels: poor (<25% percentile cut point: ≤9.999), moderately average (25% - 49% percentile cut point: 10.0 - 11.99), average (50% - 74% percentile cut point: 12.0 - 12.99), and good knowledge (≥75% percentile cut point: ≥13.0) [24]. Practices were also classified into four levels: poor (<25% percentile cut point:≤4.99), moderately average (25% - 49% percentile cut point: 5.0 - 5.999), average (50% - 74% percentile cut point: 6.0 - 6.99), and good practices (≥75% percentile cut point: ≥7.0). For all cases, the cut points were statistically evident [25,26].

Data Quality Control

To ensure the reliability and validity of the study findings, we observed the reliability analysis for both knowledge and practice variables, yielding a Cronbach α; the reliability coefficient values for the variables related to knowledge and practice were found to be 0.78 and 0.81, indicating acceptable internal consistency.

Statistical Analysis

Descriptive statistics were performed to present participants' sociodemographic characteristics and mean knowledge and practice scores. The Pearson χ^2 test and Pearson correlation coefficient were used as a bivariate analysis. Since both knowledge and practice scores did not follow normality, we performed the Mann-Whitney U test and Kruskal-Wallis 1-way ANOVA test to show the mean knowledge and practice score variations between two (eg, housewife vs working mother) and more than two groups (eg, different age groups), respectively.

Necessary assumptions were checked before performing the statistical analysis. All the data management and statistical analyses were carried out through SPSS Statistics 27.0 (IBM Corp). The *P* value was observed for all the cases at a 5% level, and 95% was considered as the CI [27-29].

Results

Sociodemographic Characteristics of the Respondents

The majority of the respondents (n=209, 52.3%) were within the 21 - 30 years age group, followed by 44% (n=176) in the 31 - 40 years age group. Most (n=57, 39.3%) respondents had a secondary level of education. Most were Muslims (n=388, 97%) and housewives (n=347, 86.8%). Many of the respondents (n=157, 39.3%) had a monthly family income of 20,001 - 40,000 BDT (US $206.19-$392.73) per month. About 13.3% (n=53) of mothers were working (Table 1).

Table . Distribution of sociodemographic characteristics of the respondents (N=400).

Characteristics	Respondents, n (%)
Age group (years)	
21 - 30	209 (52.3)
31 - 40	176 (44.0)
41 - 48	15 (3.8)
Religion	
Muslim	388 (97.0)
Non-Muslim	12 (3.0)
Educational status	
Up to primary	76 (19.0)
Secondary	157 (39.3)
Higher secondary	68 (17.0)
Bachelor's degree or higher	99 (24.8)
Occupation	
Housewife	347 (86.8)
Working	53 (13.3)
Family type	
Nuclear	272 (68.0)
Joint	128 (32.0)
Number of family members	
<5 persons	193 (48.3)
≥5 persons	207 (51.8)
Monthly family income (BDT)[a]	
≤20,000	143 (35.8)
20,001 - 40,000	157 (39.3)
≥40,001	100 (25.0)

[a]A currency exchange rate of 101.85 BDT=US $1 was used.

Knowledge Among Mothers Regarding Their Children's Oral Hygiene

Multimedia Appendix 4 shows the mothers' knowledge scores regarding their children's oral hygiene. Among the 400 mothers, more than 90% (n=360) knew the importance of brushing teeth, while 82.3% (n=329) and 80.8% (n=323) knew the recommended frequency and appropriate time for brushing teeth, respectively. Surprisingly, only 29.5% (n=118) and 38.5% (n=154) knew the duration for brushing teeth and that fluoride protects against caries, respectively. However, most of the respondents knew about the "importance of cleaning tongue" (n=365, 91.3%), "gingival disease common cause of gum bleeding" (n=286, 71.5%), "brushing and flossing protect against bleeding gum" (n=243, 60.8%), "yellow coating plaque" (n=362, 90.5%), "sugary item cause caries" (n=387, 96.8%), "soft drinks cause caries" (n=295, 73.8%), and "regular brushing protects against caries" (n=380, 95%).

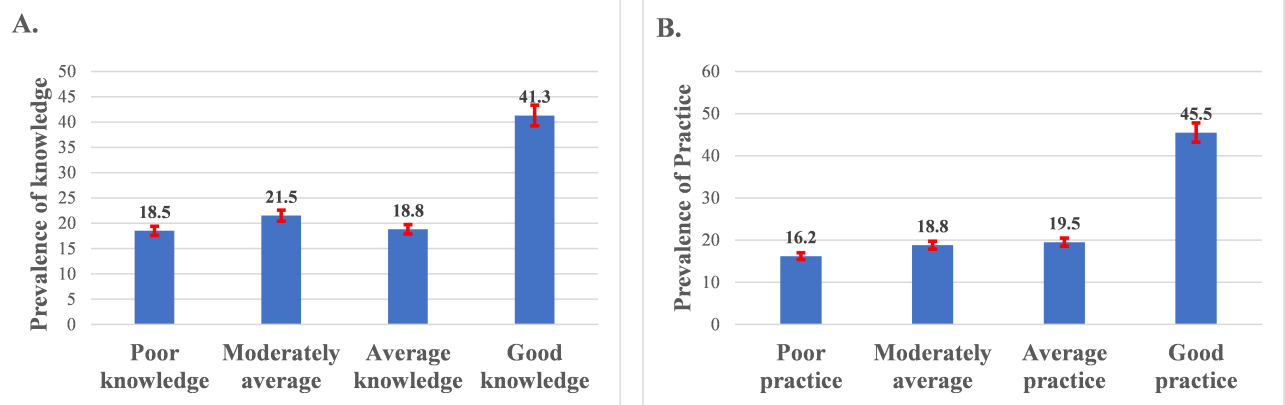
Mothers' Practices Regarding Their Children's Oral Hygiene

Multimedia Appendix 5 shows the individual distribution of mothers' practices regarding their children's oral hygiene. Most (n=381, 95.3%) of the mothers reported that their child brushed their teeth regularly, 99% (n=396) of children used a toothbrush, 62% (n=248) changed their toothbrush every 3 - 4 months or if the bristles were frayed, 97.8% (n=391) used their toothpaste, and 77.8% (n=311) rinsed their mouth after eating. Surprisingly, 44.3% (n=177) of children brushed their teeth twice daily, 42% (n=168) cleaned their tongues, and 2.8% (n=11) used floss. Only 12.5% (n=50) were given sugary items with meals, and 0.3% (n=1) were taken to dentists every 6 months.

Overall Knowledge and Practice Levels of the Respondents

Figure 1 depicts the level of knowledge and practices of mothers regarding their children's oral hygiene and the association with the mother's educational status. Only 41.3% (n=165) had good knowledge, while 18.5% (n=74) had poor knowledge (Figure 1A). Similarly, only 45.5% (n=182) of the mothers showed good practices, while 16.2% (n=65) showed poor practice levels (Figure 1B).

Figure 1. (A) Distribution of the overall knowledge of mothers. (B) Distribution of the overall practices of mothers.



Sociodemographic Variations in the Mother's Knowledge Level Regarding Their Children's Oral Hygiene

A total of 66.7% (10/15) of mothers aged 41-48 years had good knowledge regarding their children's oral hygiene. The Pearson χ^2 association test revealed that mothers' knowledge levels were significantly associated with age (P=.01), education (P<.001), family size (P=.03), and monthly income (P<.001; Table 2).

Table . Association of mothers' knowledge with sociodemographic characteristics.

Characteristics	Poor knowledge, n (%)	Moderately average, n (%)	Average knowledge, n (%)	Good knowledge, n (%)	*P* value[a]
Age group (years)					.01
21 - 30 (n=209)	43 (20.6)	55 (26.3)	39 (18.7)	72 (34.4)	
31 - 40 (n=176)	28 (15.9)	29 (16.5)	36 (20.5)	83 (47.2)	
41 - 48 (n=15)	3 (20.0)	2 (13.3)	0 (0.0)	10 (66.7)	
Religion					.22
Muslim (n=388)	72 (18.6)	86 (22.2)	72 (18.6)	158 (40.7)	
Non-Muslim (n=12)	2 (16.7)	0 (0.0)	3 (25.0)	7 (58.3)	
Educational status					<.001
Up to primary (n=76)	32 (42.1)	22 (28.9)	8 (10.5)	14 (18.4)	
Secondary (n=157)	29 (18.5)	46 (29.3)	33 (21.0)	49 (31.2)	
Higher secondary (n=68)	8 (11.8)	8 (11.8)	16 (23.5)	36 (52.9)	
Bachelor's degree or higher (n=99)	5 (5.1)	10 (10.1)	18 (18.2)	66 (66.7)	
Occupation					.10
Housewife (n=347)	68 (19.6)	77 (22.2)	67 (19.3)	135 (38.9)	
Working (n=53)	6 (11.3)	9 (17.0)	8 (15.1)	30 (56.6)	
Family type					.06
Nuclear (n=272)	46 (16.9)	52 (19.1)	59 (21.7)	115 (42.3)	
Joint (n=128)	28 (21.9)	34 (26.6)	16 (12.5)	50 (39.1)	
Number of family members					.03
<5 persons (n=193)	27 (14.0)	36 (18.7)	39 (20.2)	91 (47.2)	
≥5 persons (n=207)	47 (22.7)	50 (24.2)	36 (17.4)	74 (35.7)	
Monthly family income (BDT)[b]					<.001
≤20,000 (n=143)	38 (26.6)	37 (25.9)	25 (17.5)	43 (30.1)	
20,001 - 40,000 (n=157)	26 (16.6)	33 (21.0)	37 (23.6)	61 (38.9)	
≥41,001 (n=100)	10 (10.0)	16 (16.0)	13 (13.0)	61 (61.0)	

[a]χ^2/Fisher exact test.

[b]A currency exchange rate of 101.85 BDT=US $1 was used.

Sociodemographic Variation of the Mother's Practice Level Regarding Their Children's Oral Hygiene

Table 3 represents the association between mothers' sociodemographic characteristics and their practices regarding their children's oral hygiene. The analysis found that more than half (n=8, 53.3%) of older-aged mothers had good practices, and 66.7% (n=60) of mothers with a bachelor's degree or higher showed good practices regarding their children's oral hygiene. The educational status (*P*=.002) and income (*P*=.04) were significantly associated with the mothers' practices regarding their children's oral hygiene (Table 3).

Table . Association between sociodemographic characteristics and practice level regarding their children's oral hygiene.

Characteristics	Poor practice, n (%)	Moderately average, n (%)	Average practice, n (%)	Good practice, n (%)	*P* value[a]
Age group (years)					.34
21 - 30 (n=209)	34 (16.3)	44 (21.1)	46 (22.0)	85 (40.7)	
31 - 40 (n=176)	30 (17.0)	27 (15.3)	30 (17.0)	89 (50.6)	
41 - 48 (n=15)	1 (6.7)	4 (26.7)	2 (13.3)	8 (53.3)	
Religion of the respondents					.42
Muslim (n=388)	65 (16.8)	73 (18.8)	76 (19.6)	174 (44.8)	
Non-Muslim (n=12)	0 (0.0)	2 (16.7)	2 (16.7)	8 (66.7)	
Educational status of the respondent					.002
Up to primary (n=76)	15 (19.7)	19 (25.0)	6 (7.9)	36 (47.4)	
Secondary (n=157)	27 (17.2)	34 (21.7)	41 (26.1)	55 (35.0)	
Higher secondary (n=68)	12 (17.6)	12 (17.6)	13 (19.1)	31 (45.6)	
Bachelor's degree or higher (n=99)	11 (11.1)	10 (10.1)	18 (18.2)	60 (60.6)	
Occupation of the respondent					.24
Housewife (n=347)	60 (17.3)	68 (19.6)	65 (18.7)	154 (44.4)	
Working (n=53)	5 (9.4)	7 (13.2)	13 (24.5)	28 (52.8)	
Family type of the respondent					.98
Nuclear (n=272)	43 (15.8)	51 (18.8)	54 (19.9)	124 (45.6)	
Joint (n=128)	22 (17.2)	24 (18.8)	24 (18.8)	58 (45.3)	
Number of family members					.93
<5 persons (n=193)	30 (15.5)	38 (19.7)	36 (18.7)	89 (46.1)	
≥5 persons (n=207)	35 (16.9)	37 (17.9)	42 (20.3)	93 (44.9)	
Monthly family income (BDT)[b]					.04
≤20,000 (n=143)	30 (21.0)	30 (21.0)	28 (19.6)	55 (38.5)	
20,001 - 40,000 (n=157)	22 (14.0)	31 (19.7)	35 (22.3)	69 (43.9)	
≥41,001 (n=100)	13 (13.0)	14 (14.0)	15 (15.0)	58 (58.0)	

[a]χ^2/Fisher exact test significant level.

[b]A currency exchange rate of 101.85 BDT=US $1 is applicable.

Variation in Knowledge and Practices of the Respondents

A significant difference in respondents' knowledge and practices with sociodemographic characteristics was observed (Table 4). The analysis found that the knowledge was comparatively higher among mothers of higher age groups compared to lower age groups (mean knowledge score: 12.13, 95% CI 10.73-13.54 vs 11.23, 95% CI 10.85-11.58; *P*=.01). Similarly, both the knowledge and practice behaviors were significantly higher among mothers with higher education and income than their counterparts. In addition, working mothers and mothers with small families had significantly higher knowledge (Table 4).

Table . Knowledge and practice variation of mothers according to sociodemographic characteristics.

Characteristics	Knowledge score (range 1-15), mean (95% CI)	*P* value[a]	Practice score (range 1-13), mean (95% CI)	*P* value[a]
Age group (years)		.01		.21
21 - 30 (n=209)	11.21 (10.85 - 11.58)		6.13 (5.92 - 6.35)	
31 - 40 (n=176)	11.93 (11.56 - 12.29)		6.36 (6.09 - 6.62)	
41 - 48 (n=15)	12.13 (10.73 - 13.54)		6.8 (5.67 - 7.93)	
Religion		.22		.19
Muslim (n=388)	11.54 (11.28 - 11.80)		6.24 (6.07 - 6.41)	
Non-Muslim (n=12)	12.25 (10.53 - 13.97)		6.83 (6.08 - 7.59)	
Educational status		<.001		<.001
Up to primary (n=76)	9.66 (8.95 - 10.37)		6.01 (5.63 - 6.40)	
Secondary (n=157)	11.32 (10.97 - 11.67)		6.01 (5.75 - 6.27)	
Higher secondary (n=68)	12.26 (11.71 - 12.82)		6.19 (5.79 - 6.59)	
Bachelor's degree or higher (n=99)	12.93 (12.55 - 13.31)		6.88 (6.54 - 7.22)	
Occupation		.03		.12
Housewife (n=347)	11.45 (11.17 - 11.73)		6.21 (6.02 - 6.39)	
Working (n=53)	12.30 (11.72 - 12.89)		6.59 (6.19 - 6.98)	
Family type		.13		.88
Nuclear (n=272)	11.7 (11.39 - 12.00)		6.25 (6.05 - 6.45)	
Joint (n=128)	11.28 (10.81 - 11.75)		6.28 (5.98 - 6.59)	
Number of family members		<.001		.95
<5 persons (n=193)	11.96 (11.6 - 12.32)		6.27 (6.03 - 6.51)	
≥5 persons (n=207)	11.19 (10.84 - 11.55)		6.25 (6.01 - 6.48)	
Monthly family income (BDT)[b]		<.001		.002
≤20,000 (n=143)	10.92 (10.48 - 11.36)		5.96 (5.68 - 6.24)	
20,001 - 40,000 (n=157)	11.56 (11.17 - 11.95)		6.20 (5.96 - 6.45)	
≥40,001 (n=100)	12.49 (12.0 - 12.98)		6.77 (6.40 - 7.14)	

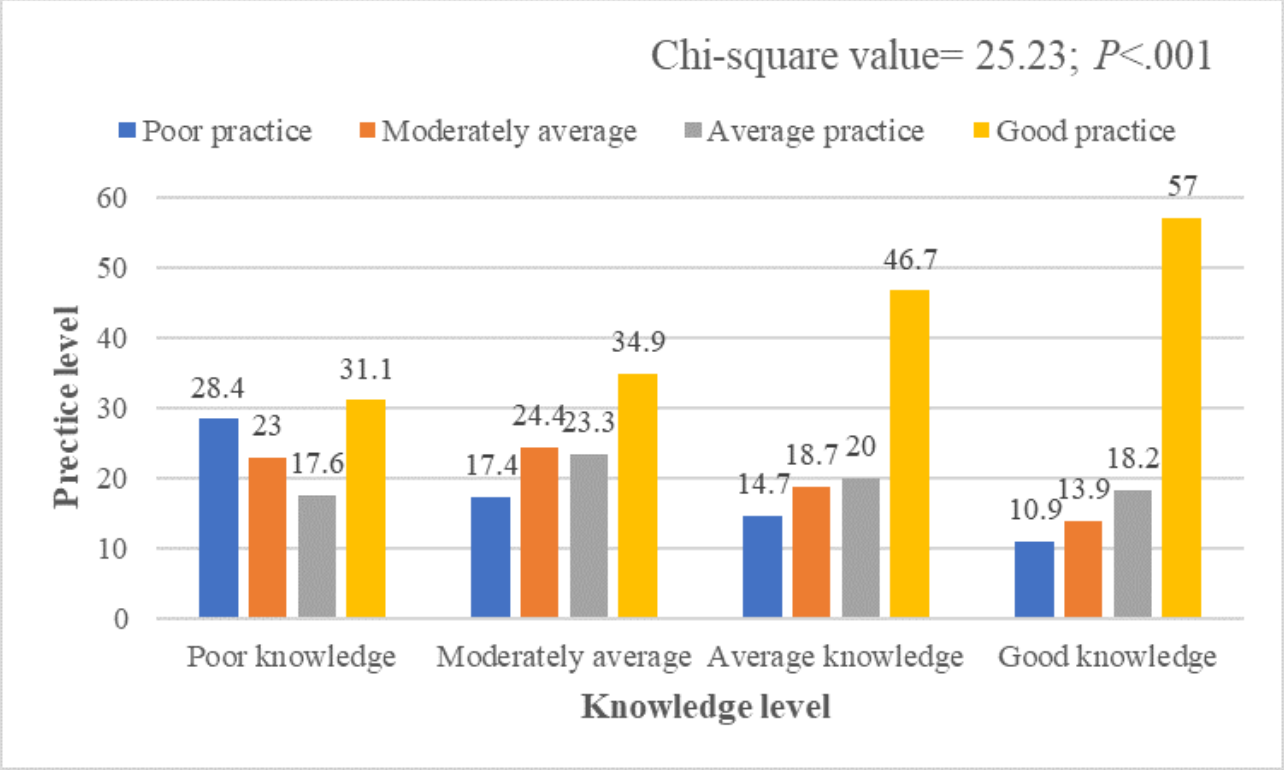
[a]Mann-Whitney *U* test and Kruskal-Wallis 1-way ANOVA test.

[b]A currency exchange rate of 101.85 BDT=US $1 is applicable.

Association Between Mothers' Oral Hygiene Knowledge and Practice Levels

Figure 2 represents the association between mothers' oral hygiene knowledge and practice levels. Over 50% of mothers with good knowledge had good practice behaviors regarding their children's oral hygiene. The Pearson correlation coefficient analysis also found a significant and positive association (r=0.301; P<.001) between the knowledge and practice scores of the respondents (Multimedia Appendix 6).

Figure 2. Association between mothers' knowledge level and practice level.



Discussion

Principal Findings

Oral health is an integral component of overall health, and it is important in our everyday lives. This study intended to evaluate mothers' knowledge and practices regarding their children's oral hygiene. An increased knowledge level was observed among older mothers, those with higher education levels, working mothers, and mothers from higher income groups. Similarly, good practices regarding children's oral hygiene were associated with the mother's education level and economic status.

Comparison to Prior Work

To maintain oral health, brushing twice a day is standard [30]. The study found that most mothers know the standard brushing recommendation for their children. Many mothers also agreed that gingival disease was the most common cause of gum bleeding, and brushing and flossing could protect against bleeding gums. The findings align with the existing literature [31]. If one wants to protect themselves against any kind of dental sickness, brushing regularly is required [32]. Over 50% of the mothers in our study agreed with this statement, which is comparable to existing research findings [33]. In this study, less than half of the mothers had good knowledge regarding their children's oral hygiene, and nearly 1 in every 5 mothers had poor knowledge. The findings suggest that health education programs among mothers regarding their children's oral hygiene are needed. Various education and awareness programs, including television, social and mass media campaigns, and community-based educational interventions may improve mothers' knowledge regarding children's oral hygiene [34-36].

In this study, the mother's knowledge regarding their children's oral hygiene was significantly associated with their age, and mothers in the higher age group had comparatively higher knowledge than those in the lower age group. The finding is comparable to many studies that suggest oral health educational programs for younger mothers [34,37,38]. The mother's educational status and monthly family income were two important predictors for increasing their children's oral hygiene knowledge and practices. Parents with higher education were more aware of their children's dental health [39,40]. Our research results align with the existing literature that indicates that mothers who have attained a university degree possess superior knowledge about oral health in comparison to those with a lower level of education [41]. This might be rationalized by the deduction that women with a lower level of education may lack awareness about the consequences of probable risk factors linked to the progression of oral disorders. Consequently, health awareness and promotion play a vital role for mothers who have inadequate educational backgrounds [40,42,43]. Our results align with the existing research, which demonstrates that mothers with extensive knowledge tend to promote good oral health habits in their children [25].

Strengths and Limitations

This study aimed to identify the variables that impact oral hygiene habits among mothers and evaluate their level of knowledge and compliance with oral hygiene practices. The primary merit of this study is the results. We identified the variables that influence individuals' understanding and behaviors related to oral hygiene. We experienced a few limitations during this study. First, this was cross-sectional research, which lacks strength in cause-effect analysis. Second, the study was conducted among mothers visiting tertiary-level hospitals in

Dhaka. Therefore, there is a chance of nonresponse bias due to convenience sampling.

Future Directions

Maintaining good oral hygiene is crucial for every child's overall health; mothers, in particular, play a vital role in this regard. Based on our study findings, the following recommendations may help enhance maternal knowledge and improve children's oral hygiene practices.

Educational Workshops and School-Based Initiatives

Community-based educational programs including workshops and seminars may help educate mothers of different age groups [34,40]. These workshops should focus on the importance of oral hygiene, practical tips for maintaining children's oral health, and common misconceptions. Monthly informational sessions on oral hygiene practices facilitated by dental health professionals and community health centers could play an important role in improving children's oral hygiene practices. Various school-based initiatives, like partnering with schools to offer regular seminars and distributing informative materials to parents during parent-teacher meetings that emphasize the critical role of oral hygiene from an early age, could be implemented [37].

Incorporate Oral Health Education Into the Curriculum

Integration of basic oral health education into the curriculum of early childhood education programs, ensuring that children learn about oral hygiene from a young age, may help children improve their oral hygiene practices [39,44]. Various programs within schools that encourage parental involvement in learning about and practicing good oral hygiene, and providing resources and support for mothers to reinforce these practices at home may help children improve their oral hygiene practices [44].

Media and Technology Use

Launching social media campaigns targeting mothers; using platforms like Facebook, Instagram, and YouTube to disseminate information on children's oral hygiene; and featuring engaging content such as infographics, videos, and interactive question-and-answer sessions with dental professionals could also be influential initiatives [35].

Research and Monitoring

Support should also be provided for ongoing research to monitor the effectiveness of these initiatives and to identify new trends and needs related to children's oral hygiene [45]. Establishing feedback mechanisms, such as surveys and focus groups, can help gather insights from mothers on the effectiveness of current programs and identify areas for improvement.

Conclusion

This study revealed that mothers' knowledge and practices regarding their children's oral health were insufficient. The mother's age, education level, family size, and monthly income significantly influenced their knowledge level. Children's oral hygiene habits were significantly associated with family income and the mother's educational status. Women aged 41-48 years with a bachelor's degree or higher, from higher socioeconomic backgrounds, and with school-aged children demonstrated significantly higher levels of knowledge. Mothers with higher socioeconomic status and more education demonstrated a much higher level of dental hygiene practices for their children. The mother's knowledge regarding their children's oral hygiene had positive effects and significantly correlated with their children's oral hygiene practices. The findings of this study emphasize the need for educational and school-based initiatives, accessible dental care services, oral health education in the curriculum, media and technology involvement in oral health educational campaigns, and proper research and monitoring.

Acknowledgments

We acknowledge the Department of Biostatistics, National Institute of Preventive and Social Medicine for their technical support during the study. We are also grateful to all participants included in this study. We are thankful to the Supporo Dental College and Hospital for their administrative support during data collection.

Data Availability

The datasets generated or analyzed during this study were deposited onto figshare [46].

Authors' Contributions

Conceptualization: TT, MMR, HS
Formal analysis: TT, MMR
Investigation: TT, SKD, AN, FN, NN, THB, SAS, SKK, MABS, SMR, UH, ZF, MMR
Methodology: TT, MMR
Project administration: TT, MMR
Supervision: MMR
Funding acquisition: TT, SKD, FN, NN, THB, SAS, SKK, SMR, UH, ZF, AAK, MMR
Validation: HS, AAK, MMR
Visualization: TT, SKD, MMR
Writing - original draft: TT, SKD, MMR
Investigation: TT, SKD, AN, FN, NN, THB, SAS, SKK, MABS, SMR, UH, ZF, HS, AAK, MMR

Writing - review and editing: TT, SKD, AN, FN, NN, THB, SAS, SKK, MABS, SMR, UH, ZF, HS, AAK, MMR

Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement: checklist of items that should be included in reports of cross-sectional studies.
[PDF File, 76 KB - xmed_v6i1e59379_app1.pdf]

Multimedia Appendix 2
List of variables used to assess mothers' knowledge regarding their children's oral hygiene.
[DOCX File, 14 KB - xmed_v6i1e59379_app2.docx]

Multimedia Appendix 3
List of variables used to assess mothers' practices regarding their children's oral hygiene.
[DOCX File, 14 KB - xmed_v6i1e59379_app3.docx]

Multimedia Appendix 4
Mothers' individual knowledge regarding their children's oral hygiene.
[DOCX File, 14 KB - xmed_v6i1e59379_app4.docx]

Multimedia Appendix 5
Mothers' individual practices regarding their children's oral hygiene.
[DOCX File, 14 KB - xmed_v6i1e59379_app5.docx]

Multimedia Appendix 6
Correlation between knowledge and practice scores.
[DOCX File, 13 KB - xmed_v6i1e59379_app6.docx]

References

1. Global oral health status report: towards universal health coverage for oral health by 2030. World Health Organization. 2022 Nov 18. URL: https://www.who.int/publications/i/item/9789240061484 [accessed 2025-01-13]
2. Alama MA, Abdullah US, Mofiz M, Aktar S, Karim M. Oral health status among the under five children attending at OPD of Dhaka Dental College Hospital. Update Dent Coll J 2016 Apr 7;5(2):9-17. [doi: 10.3329/updcj.v5i2.27263]
3. Haque MF, Rahman MM, Alif SM, et al. Estimation and prediction of doubling time for COVID-19 epidemic in Bangladesh: a study of first 14 month's daily confirmed new cases and deaths. Global Biosecurity 2021 Apr 26;3(1). [doi: 10.31646/gbio.91]
4. Shah PM, Jeevanadan G. Oral hygiene maintenance in children - a survey on parental awareness. Int J Pharm Res 2017;9(3). [doi: 10.31838/ijpr/2020.12.01.311]
5. Toothbrushes. American Dental Association. 2019. URL: https://www.ada.org/resources/research/science-and-research-institute/oral-health-topics/toothbrushes [accessed 2022-11-27]
6. Lawal FB, Fagbule OF, Akinloye SJ, Lawal TA, Oke GA. Impact of oral hygiene habits on oral health-related quality of life of in-school adolescents in Ibadan, Nigeria. Front Oral Health 2022 Sep 9;3:979674. [doi: 10.3389/froh.2022.979674] [Medline: 36338573]
7. Saadaldina SA, Eldwakhly E, Alnazzawi AA, et al. Awareness and practice of oral health measures in Medina, Saudi Arabia: an observational study. Int J Environ Res Public Health 2020 Dec 6;17(23):1-10. [doi: 10.3390/ijerph17239112] [Medline: 33291281]
8. Oral health. World Health Organization. 2023. URL: https://www.who.int/health-topics/oral-health#tab=tab_1 [accessed 2025-01-20]
9. Duangthip D, Chu CH. Challenges in oral hygiene and oral health policy. Front Oral Health 2020 Oct 7;1:575428. [doi: 10.3389/froh.2020.575428] [Medline: 35047981]
10. Anil S, Anand PS. Early childhood caries: prevalence, risk factors, and prevention. Front Pediatr 2017 Jul 18;5:157. [doi: 10.3389/fped.2017.00157] [Medline: 28770188]
11. Hossain MS, Tasnim S, Chowdhury MA, Chowdhury FIF, Hossain D, Rahman MM. Under‐five children's acute respiratory infection dropped significantly in Bangladesh: an evidence from Bangladesh demographic and health survey, 1996–2018. Acta Paediatr 2022 Oct;111(10):1981-1994. [doi: 10.1111/apa.16447] [Medline: 35678484]

12. Islam GMR, Rahman MM, Hasan MI, Tadesse AW, Hamadani JD, Hamer DH. Hair, serum and urine chromium levels in children with cognitive defects: a systematic review and meta-analysis of case control studies. Chemosphere 2022 Mar;291(Pt 2):133017. [doi: 10.1016/j.chemosphere.2021.133017] [Medline: 34813844]
13. Sony SA, Haseen F, Islam SS, Chowdhury SF. Knowledge and practice of oral health and hygiene and oral health status among school going adolescents in a rural area of Sylhet District. Community Based Med J 2022 Jan 10;10(1):30-36. [doi: 10.3329/cbmj.v10i1.58642]
14. Mohammadi TM, Hajizamani A, Bozorgmehr E. Improving oral health status of preschool children using motivational interviewing method. Dent Res J (Isfahan) 2015;12(5):476-481. [doi: 10.4103/1735-3327.166231] [Medline: 26604963]
15. Holm AK. Caries in the preschool child: international trends. J Dent 1990 Dec;18(6):291-295. [doi: 10.1016/0300-5712(90)90125-x] [Medline: 2074302]
16. Al-Batayneh OB, Al-Khateeb HO, Ibrahim WM, Khader YS. Parental knowledge and acceptance of different treatment options for primary teeth provided by dental practitioners. Front Public Health 2019 Nov 7;7:322. [doi: 10.3389/fpubh.2019.00322] [Medline: 31788466]
17. Chand S, Chand S, Dhanker K, Chaudhary A. Impact of mothers′ oral hygiene knowledge and practice on oral hygiene status of their 12-year-old children: a cross-sectional study. J Indian Assoc Public Health Dent 2014;12(4):323-329. [doi: 10.4103/2319-5932.147681]
18. Khodadadi E, Niknahad A, Sistani MMN, Motallebnejad M. Parents' oral health literacy and its impact on their children's dental health status. Electron Physician 2016 Dec 25;8(12):3421-3425. [doi: 10.19082/3421] [Medline: 28163858]
19. Al-Zahrani AM, Al-Mushayt AS, Otaibi MF, Wyne AH. Knowledge and attitude of Saudi mothers towards their preschool children's oral health. Pak J Med Sci 2014 Jul;30(4):720-724. [doi: 10.12669/pjms.304.5069] [Medline: 25097504]
20. Mohandass B, Chaudhary H, Pal GK, Kaur S. Knowledge and practice of rural mothers on oral hygiene for children. Indian J Continuing Nurs Education 2021;22(1):39-43. [doi: 10.4103/IJCN.IJCN_7_20]
21. Ali Leghari M, Tanwir F, Ali H. Association of dental caries and parents knowledge of oral health, a cross-sectional survey of schools of Karachi, Pakistan. J Pakistan Dent Assoc 2018 May 15;23(1) [FREE Full text]
22. Bennadi D, Reddy CVK, Sunitha S, Kshetrimayum N. Oral health status of 3-6 year old children and their mother's oral health related knowledge, attitude and practices in Mysore City, India. Asian J Med Sci 2014 Sep 15;6(2):66-71. [doi: 10.3126/ajms.v6i2.11097]
23. Bakar N, Mamat Z. Parental knowledge and practices on preschool children oral healthcare in Nibong Tebal Penang, Malaysia. JOJ Nurs Health Care 2018 Apr;7(4). [doi: 10.19080/JOJNHC.2018.07.555716]
24. Kabar AME, Elzahaf RA, Shakhatreh FM. The relationship between oral health knowledge mothers and dental caries in Tripoli, Libya. Saudi J Oral Dent Res 2019 Jul 22;4(7):463-467. [doi: 10.21276/sjodr.2019.4.7.7]
25. Alzaidi SS, Alanazi IA, Abo Nawas OM, Mulla MA. Childhood oral health: maternal knowledge and practice in Tabuk, Saudi Arabia. Egyptian J Hosp Med 2018 Jan;70(9):1544-1551. [doi: 10.12816/0044681]
26. Das SK, Tamannur T, Nesa A, et al. Exploring the knowledge and practices on road safety measures among motorbikers in Dhaka, Bangladesh: a cross-sectional study. Inj Prev 2023 Nov 28:ip-2023-045071. [doi: 10.1136/ip-2023-045071] [Medline: 38050086]
27. Nachar N. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. Tutorials Quant Methods Psychol 2008;4(1):13-20. [doi: 10.20982/tqmp.04.1.p013]
28. Kim HY. Statistical notes for clinical researchers: nonparametric statistical methods: 2. Nonparametric methods for comparing three or more groups and repeated measures. Restor Dent Endod 2014 Nov;39(4):329-332. [doi: 10.5395/rde.2014.39.4.329] [Medline: 25383354]
29. Rahman MM, Hamiduzzaman M, Akter MS, et al. Frailty indexed classification of Bangladeshi older adults' physio-psychosocial health and associated risk factors- a cross-sectional survey study. BMC Geriatr 2021 Jan 6;21(1):3. [doi: 10.1186/s12877-020-01970-5] [Medline: 33402094]
30. Manzoor F, Iqbal Z, Ahmed K, Khayyam U, Malhi P, Khalid M. Assessment of parental knowledge and attitude regarding oral health status of their children in District Mirpurkhas Sindh, Pakistan. Pakistan J Med Health Sci 2021 Apr;15(4):1352-1355 [FREE Full text]
31. Salama AA, Konsowa EM, Alkalash SH. Mothers' knowledge, attitude, and practice regarding their primary school children's oral hygiene. Menoufia Med J 2020 Mar 25;33(1):11-17. [doi: 10.4103/mmj.mmj_300_19]
32. Tonetti MS, Bottenberg P, Conrads G, et al. Dental caries and periodontal diseases in the ageing population: call to action to protect and enhance oral health and well-being as an essential component of healthy ageing - consensus report of group 4 of the joint EFP/ORCA workshop on the boundaries between caries and periodontal diseases. J Clin Periodontol 2017 Mar;44 Suppl 18:S135-S144. [doi: 10.1111/jcpe.12681] [Medline: 28266112]
33. Ibrahim R, Helaly MO, Ahmed EMA. Assessment of brushing techniques in school children and its association with dental caries, Omdurman, 2019. Int J Dent 2021 Jan 21;2021:4383418. [doi: 10.1155/2021/4383418] [Medline: 33552159]
34. Amin M, Nyachhyon P, Elyasi M, Al-Nuaimi M. Impact of an oral health education workshop on parents' oral health knowledge, attitude, and perceived behavioral control among African immigrants. J Oral Dis 2014 Jun 23;2014:1-7. [doi: 10.1155/2014/986745]

35. Sharma S, Mohanty V, Balappanavar AY, Chahar P, Rijhwani K. Role of digital media in promoting oral health: a systematic review. Cureus 2022 Sep 7;14(9):e28893. [doi: 10.7759/cureus.28893] [Medline: 36225421]
36. Goldberg E, Eberhard J, Bauman A, Smith BJ. Mass media campaigns for the promotion of oral health: a scoping review. BMC Oral Health 2022 May 14;22(1):182. [doi: 10.1186/s12903-022-02212-3] [Medline: 35568896]
37. Suresh BS, Ravishankar TL, Chaitra TR, Mohapatra AK, Gupta V. Mother's knowledge about pre-school child's oral health. J Indian Soc Pedod Prev Dent 2010;28(4):282-287. [doi: 10.4103/0970-4388.76159] [Medline: 21273717]
38. Alshammari FS, Alshammari RA, Alshammari MH, et al. Parental awareness and knowledge toward their children's oral health in the city of Dammam, Saudi Arabia. Int J Clin Pediatr Dent 2021;14(1):100-103. [doi: 10.5005/jp-journals-10005-1894] [Medline: 34326593]
39. Jumaa FA, Turki SG, Hattab KM. Mothers' knowledge toward oral health of children under 5 years old. Pakistan J Med Health Sci 2022;16(6):437-442. [doi: 10.53350/pjmhs22166437]
40. Nepaul P, Mahomed O. Influence of parents' oral health knowledge and attitudes on oral health practices of children (5-12 years) in a rural school in KwaZulu-Natal, South Africa. J Int Soc Prev Community Dent 2020 Sep 28;10(5):605-612. [doi: 10.4103/jispcd.JISPCD_273_20] [Medline: 33282770]
41. Sehrawat P, Shivlingesh KK, Gupta B, Anand R, Sharma A, Chaudhry M. Oral health knowledge, awareness and associated practices of pre-school children's mothers in Greater Noida, India. Niger Postgrad Med J 2016;23(3):152-157. [doi: 10.4103/1117-1936.190344] [Medline: 27623728]
42. Đorđević A. Parents' knowledge about the effects of oral hygiene, proper nutrition and fluoride prophylaxis on oral health in early childhood. Balkan J Dent Med 2018;22(3):26-31. [doi: 10.2478/bjdm-2018-0005]
43. Gurunathan D, Moses J, Arunachalam SK. Knowledge, attitude, and practice of mothers regarding oral hygiene of primary school children in Chennai, Tamil Nadu, India. Int J Clin Pediatr Dent 2018;11(4):338-343. [doi: 10.5005/jp-journals-10005-1535] [Medline: 30397379]
44. Das H, Janakiram C, Ramanarayanan V, et al. Effectiveness of an oral health curriculum in reducing dental caries increment and improving oral hygiene behaviour among schoolchildren of Ernakulam district in Kerala, India: study protocol for a cluster randomised trial. BMJ Open 2023 Feb 20;13(2):e069877. [doi: 10.1136/bmjopen-2022-069877] [Medline: 36806129]
45. Chawłowska E, Karasiewicz M, Lipiak A, et al. Exploring the relationships between children's oral health and parents' oral health knowledge, literacy, behaviours and adherence to recommendations: a cross-sectional survey. Int J Environ Res Public Health 2022 Sep 8;19(18):11288. [doi: 10.3390/ijerph191811288] [Medline: 36141563]
46. Rahman MM. Knowledge and practices towards oral hygiene of children aged 5-9 years old: a cross-sectional dataset among mothers visited tertiary level hospitals. figshare. 2024 Aug 31. URL: https://doi.org/10.6084/m9.figshare.26886547.v1 [accessed 2024-01-13]

Abbreviations

ECC: early childhood caries
NIPSOM: National Institute of Preventive and Social Medicine
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by E Meinert, T Leung; submitted 10.04.24; peer-reviewed by B Nwankwo, MH Islam; revised version received 13.09.24; accepted 27.11.24; published 03.02.25.

Please cite as:
Tamannur T, Das SK, Nesa A, Nahar F, Nowshin N, Binty TH, Shakil SA, Kundu SK, Siddik MAB, Rafsun SM, Habiba U, Farhana Z, Sultana H, Kamil AA, Rahman MM
Mothers' Knowledge of and Practices Toward Oral Hygiene of Children Aged 5-9 Years in Bangladesh: Cross-Sectional Study
JMIRx Med 2025;6:e59379
URL: https://xmed.jmir.org/2025/1/e59379
doi:10.2196/59379

Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis

Bernard Friedenson, PhD

Department of Biochemistry and Medical Genetics, Cancer Center, University of Illinois Chicago, 900 s Ashland, Chicago, IL, United States

Corresponding Author:
Bernard Friedenson, PhD
Department of Biochemistry and Medical Genetics, Cancer Center, University of Illinois Chicago, 900 s Ashland, Chicago, IL, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.07.03.23292185v1

Companion article: https://med.jmirx.org/2025/1/e70039

Companion article: https://med.jmirx.org/2025/1/e70041

Companion article: https://med.jmirx.org/2025/1/e69307

Abstract

Background: The causes of breast cancer are poorly understood. A potential risk factor is Epstein-Barr virus (EBV), a lifelong infection nearly everyone acquires. EBV-transformed human mammary cells accelerate breast cancer when transplanted into immunosuppressed mice, but the virus can disappear as malignant cells reproduce. If this model applies to human breast cancers, then they should have genome damage characteristic of EBV infection.

Objective: This study tests the hypothesis that EBV infection predisposes one to breast cancer by causing permanent genome damage that compromises cancer safeguards.

Methods: Publicly available genome data from approximately 2100 breast cancers and 25 ovarian cancers were compared to cancers with proven associations to EBV, including 70 nasopharyngeal cancers, 90 Burkitt lymphomas, 88 diffuse large B-cell lymphomas, and 34 gastric cancers. Calculation algorithms to make these comparisons were developed.

Results: Chromosome breakpoints in breast and ovarian cancer clustered around breakpoints in EBV-associated cancers. Breakpoint distributions in breast and EBV-associated cancers on some chromosomes were not confidently distinguished ($P>.05$), but differed from controls unrelated to EBV infection. Viral breakpoint clusters occurred in high-risk, sporadic, and other breast cancer subgroups. Breakpoint clusters disrupted gene functions essential for cancer protection, which remain compromised even if EBV infection disappears. As CRISPR (clustered regularly interspaced short palindromic repeats)–like reminders of past infection during evolution, EBV genome fragments were found regularly interspaced between Piwi-interacting RNA (piRNA) genes on chromosome 6. Both breast and EBV-associated cancers had inactivated genes that guard piRNA defenses and the major histocompatibility complex (MHC) locus. Breast and EBV-associated cancer breakpoints and other variations converged around the highly polymorphic MHC. Not everyone develops cancer because MHC differences produce differing responses to EBV infection. Chromosome shattering and mutation hot spots in breast cancers preferentially occurred at incorporated viral sequences. On chromosome 17, breast cancer breakpoints that clustered around those in EBV-mediated cancers were linked to estrogen effects. Other breast cancer breaks affected sites where EBV inhibits JAK-STAT and SWI-SNF signaling pathways. A characteristic EBV-cancer gene deletion that shifts metabolism to favor tumors was also found in breast cancers. These changes push breast cancer into metastasis and then favor survival of metastatic cells.

Conclusions: EBV infection predisposes one to breast cancer and metastasis, even if the virus disappears. Identifying this pathogenic viral damage may improve screening, treatment, and prevention. Immunizing children against EBV may protect against breast, ovarian, other cancers, and potentially even chronic unexplained diseases.

(JMIRx Med 2025;6:e50712) doi:10.2196/50712

KEYWORDS

breast cancer; cancer; oncology; ovarian; virus; viral; Epstein-Barr; herpes; bioinformatics; chromosome; gene; genetic; chromosomal; DNA; genomic; BRCA; metastasis; biology

Introduction

In the United States, over 40,000 women die from breast cancer each year [1,2]. The causes of the disease are not well understood, making prevention and treatment empirical and hazardous. At the time of breast cancer diagnosis, its causes are difficult to isolate from multiple risk factors. A human cancer virus is one such risk factor. A tumor virus does not cause cancer by itself [3] but can make cancer more likely by inhibiting tumor suppressors [4] or activating oncogenes. Viral damage then increases cancer risks via mutations and chromosome breaks. Epstein-Barr virus (EBV), also called human herpesvirus 4, infects at least 90% of humans as a lifelong infection, often acquired at an early age [5], but the virus remains latent and asymptomatic in most people. EBV may be a risk factor for breast cancer. Active infection is significantly more prevalent in breast cancer tissues than in normal and benign controls [6], increasing risk by 4.75- to 6.29-fold [7]. EBV transformed human mammary epithelial cells in culture so that xenografts in immunosuppressed mice accelerated breast cancer. Once malignant transformation occurred, EBV was no longer required [8], but the cells remain malignant.

There has been no way to test the idea that EBV causes breast cancer and can then disappear. However, cancers in other tissues have proven relationships to EBV infection, so these known EBV-associated cancers can be compared to breast cancers at the genome level. Cancers with unambiguous EBV associations include nasopharyngeal cancer (NPC), EBV-positive diffuse large B-cell lymphoma (DLBCL), endemic Burkitt lymphoma (BL) [9], and gastric cancer (GC). Some genomic similarities between these EBV-associated cancers and breast cancer can be derived from the literature. In NPC, 100% of malignant cells are EBV positive [10]. Over 64% of NPCs are deficient in a pathway that depends on the breast cancer susceptibility genes *BRCA1* and *BRCA2* [11], which accurately repair DNA crosslinks and breaks via the homologous recombination pathway. This sprawling, interconnected pathway includes Fanconi anemia (FA) gene products and is often designated as the FA-BRCA pathway. In 126 patients with NPC, *BRCA1* and *BRCA2* were the most frequently mutated genes (55.5% and 33.3%, respectively) [12]. NPC mutations interfere with innate immunity and constitutively activate an inflammatory response. Overexpressed nuclear factor–κB (NF-κB) is a hallmark of NPC, occurring in 90% of NPCs [11]. Similarly, almost all stage-3 breast cancers overexpress NF-κB [13].

In NPC and the other known EBV-associated cancers, EBV inhibits the FA-BRCA pathway by various methods, including using viral microRNAs to downregulate *BRCA1* [14], hijacking other pathway components [15,16], and destabilizing SMC5/6-mediated chromatin interactions [17,18]. In GC, EBV infection and FA-BRCA pathway status are mutually exclusive [19], implying that EBV infection is approximately equivalent to disabling the FA-BRCA pathway. In DLBCL, the best prognostic marker is FA-BRCA pathway status [20]. In DLBCL and endemic BL, EBV variant infection accompanies *MYC* translocations. These translocations drive the disease and make a characteristic replacement of normal *MYC* control elements with highly active immunoglobulin regulatory sequences [21,22]. *MYC* amplification is frequent in breast cancers that have inactive *BRCA1* [23].

NPCs, DLBCLs, BLs, GCs, and breast cancers all have deficits in correctly repairing double-strand breaks and crosslinks. The compromised FA-BRCA pathway can produce chromosomes with too many centromeres. During cell division, mitotic spindles pull chromatids with multiple centromeres in too many directions, generating chromosome breaks to destabilize the human genome [24,25]. In breast cancer, these variations mark breakpoints at translocations and oncogene amplifications [26].

If EBV contributes to breast cancer, gene deficits in breast cancers and EBV-associated cancers should produce comparable changes in the human genome that do not depend on whether EBV infection persists. The aim of this study was to test for these virus-induced genome changes using bioinformatic calculations and analyses. The results could implicate EBV and its variants in disabling a variety of molecular and cellular safeguards that protect against breast cancer and its metastasis. Whether or not cancer develops in response to EBV infection depends on major histocompatibility complex (MHC) gene polymorphisms [27,28], so not everyone infected with EBV will develop cancer. In susceptible people, genome damage is permanent and does not require large numbers of viral particles, active infection, or continuing virus presence. Childhood immunization against selected EBV gene products may do much to prevent breast, ovarian, and other cancers.

Methods

Datasets Used in the Analysis

Overview

The initial data for analysis came from literature searches for studies on breast and EBV-associated cancers with large numbers of participants, unrestricted access to genome information, and complete whole-genome analysis. The first criterion for including breast cancer data was published intrachain or interchain chromosome breakpoints from high-quality, peer-reviewed publications produced by world-class laboratories. The second criterion was the availability of sufficient DNA sequence data to specify the location of these chromosome breakpoints. The third criterion was that genome sequencing had been done on samples taken before treatment began. These publicly available DNA sequence data were chosen to encompass diverse genetics, subtypes, stages, grades, morphologies, and outcomes. Initially, breast cancers were separated only broadly into those with a likely hereditary component versus those without this component. The cancers had to include typical morphologies such as ductal carcinomas, lobular carcinomas, medullary carcinomas, and

invasive carcinomas (ie, “no special type”). The included breast cancers were all primary stage-2 or stage-3 cancers. Although surgery usually removes these primary tumors, cells with only a few additional late mutations are responsible for seeding local recurrences or metastases, so primary and metastatic tumors are not very different [29]. Although the selected cancers are not a random sample representing all breast cancers [30], they are likely to have chromosome instability originating from diverse typical causes.

Specifically, the breast cancer data used came from 560 breast cancer genome sequences, familial cancer data from 78 patients, methylation data from 1538 breast cancers versus 244 controls, 243 triple-negative breast cancers, and 2658 human cancers [31-35]. Data also included 74 breast cancers from high-risk women who were typed as having *BRCA1*- or *BRCA2*-associated mutations or cancers diagnosed before the age of 40 years [36,37]. Another study of familial breast cancers contributed 65 familial breast cancers [33]. Gene breakpoints for many interchromosomal and intrachromosomal translocations and breakpoints were obtained from the COSMIC (Catalog of Somatic Mutations in Cancer) website, as curated from original publications or original articles and their supplemental informationsupplemental information [31-33]. Multimedia Appendix 1 provides a glossary of the terms used in this paper.

Breakpoints in Breast Cancers From High-Risk Women

Hereditary cancers were taken as breast cancers from women with a typed high-risk *BRCA1* or *BRCA2* mutation diagnosed before the age of 70 years. Cancers from patients with onset before the age of 50 years were also included to add more data, since these women are at high risk for an inherited, cancer-associated mutation. These patient samples were chosen based on descriptions in published data defining the breast cancer cohorts [31,33].

Sporadic Breast Cancers

Sporadic breast cancers were taken as breast cancers diagnosed after the age of 70 years that did not have a known inherited mutation [31].

Breast Cancer Subgroups

Human epidermal growth factor receptor 2 (HER2)–positive and triple-negative breast cancer data used for subgroup analysis were from original publications [33] and the COSMIC website.

Exclusions

Male breast cancers were excluded.

Data Source for Ovarian Cancers

Data for breakpoints in ovarian cancers were downloaded from the COSMIC website. The cancers corresponded to “mixed adenosquamous ovarian carcinomas” and were arbitrarily taken from those with the largest number of structural variants. These cancers all had the prefix “AOCS-” with further identification numbers and *BRCA* mutation status in parentheses as follows: 170-1-8 (negative), 120-3-6 (*BRCA2*), 142-3-5 (negative), 139-1-5 (negative), 086-3-2 (negative), 147-1-1 (*BRCA1* and *BRCA2*), 094 - 6-X (*BRCA1*), 094-1-1 (*BRCA1*), 088-3-8 (negative), 139-6-3 (*BRCA2*), 150-3-1 (negative), 116-1-3 (negative), 155-3-5 (*BRCA2*), 093-3-6 (negative), 034-3-8 (*BRCA1*), 091-3-0 (*BRCA1*), 139-19-0 (*BRCA2*), 170-3-5 (negative), 114-1-8 (negative), 064-3-3 (negative), 064-1-6 (negative), 106-1-1 (*BRCA1*), 152 - 1-X (*BRCA1*), and 134-1-5 (unknown).

Original Data Sources for Cancers With Known EBV Associations: NPCs, Lymphomas, and GCs

Overview

NPC chromosome breakpoint positions were retrieved from Bruce et al [11] for 70 primary tumors of the nasopharynx at stages 1-4C. The data came from whole-genome sequencing of “63 micro-dissected tumors, 5-patient derived xenografts, and two cell lines.” DLBCL breakpoints were collected from 88 patients with DLBCL (aged >60 y) [22]. The *MYC* breakpoints included class I and II *MYC* translocation locus breakpoints defined in BL, encompassing areas far upstream of *c-myc* [38-40]. Downstream breakpoints included an enhancer region approximately 565 kilobases long on the nearest telomere side of the *MYC* coding sequence [22]. Older data provided fusion sequences as Gencode Accession numbers [21]. These fusion sequences were downloaded as FASTA files and copied to BLAST (Basic Local Alignment Search Tool) for placement on the human GRCh38/hg38 reference sequence. GCs with inferred EBV infection status came from 34 (20.2%) out of 168 samples subjected to whole-genome sequencing [41].

Selection Bias

As much as possible, selection bias was avoided by blindly selecting samples, replicating samples using cohorts from different publications, using the largest possible groups of samples, and avoiding convenience sampling. Some experiments used a newer dataset from 780 breast cancers [22] for comparisons to confirm that selection bias was unlikely.

Recruitment

Data from genome sequence studies did not include specific recruitment procedures for patients with cancer. However, patients are typically recruited through hospitals and clinics with referrals from medical professionals. Patients provide informed consent to have their genomes sequenced and used for research and to integrate cancer genome sequence data into treatment decisions [42].

Methods Used to Determine That DNA Breakpoints From Breast and Ovarian Cancers Clustered Around Breakpoints in EBV-Associated Cancers

Calculation of Distances Between Breakpoints in Breast and Ovarian Cancers Versus EBV-Related Cancers

Before combining or comparing datasets, they were all converted to the same genome version, usually GRCh38. The break position in breast cancer nearest to a break in NPC was taken as the Microsoft Excel *XLOOKUP* value for the number of base pairs (bp) from the closest NPC breakpoint 5’ to the breast cancer break or the NPC breakpoint 3’ to the breast cancer break, whichever was closer (Multimedia Appendix 2). For comparing a given breast cancer breakpoint A2 to EBV-associated cancer breakpoints B2 to B72, the initial

algorithm to find the nearest 5' break position was written in cell C2 as follows:*=XLOOKUP($A2,$B$2:$B$72,$B$2:$B$72,0, −1,1)*. Changing −1 to +1 gave D2, the nearest 3' position. Distance from the breast cancer breakpoint was then calculated as *=MIN(ABS(C2-A2), ABS(D2-A2))*. The same formulas were then continuously updated by Excel to calculate all other breast cancer comparisons in column A. Differences in the amount of data available for NPC versus breast cancer breakpoints complicated the calculations near chromosome telomeres. Several methods of handling these end regions made no discernible difference in the outcomes. For a 5000-bp window, an overflow window of 5,000,000 was used to limit the number of bins to a maximum of 1000. Another method of calculating distances between chromosome breakpoints in different cancers used the minimum of the absolute values of distances between breast cancers and the array of breakpoints in GCs, BLs, or NPCs. This method gave results identical to *XLOOKUP* values but was more convenient to compare clusters of breast cancer breakpoints to those in lymphoid and epithelial EBV-associated cancers. Hundreds of millions of calculations were repeated at least twice. Most of the calculations in this section are presented in Multimedia Appendix 2.

DNA Sequence Homology Analyses to Determine Breakpoints in Human Cancer Sequences That Resemble Viral Sequences

The NCBI BLASTn program (MegaBLAST) and database [43-45] were used to compare DNA sequence homologies around breakpoints in breast cancers to all available viral DNA sequences. *E* ("expect") values are related to *P* values and represent the probability that a given homology bit score occurs by chance. *E* values $<1\times10^{-10}$ were considered significant homology. In many cases, *E* values were "0" ($<1\times10^{-180}$) and always far below 1×10^{-10}. The virus DNA was retrieved from BLAST searches using "viruses (taxid:10239)," with human sequences, mouse sequences, and uncharacterized sample mixtures excluded. Different strains and isolates of the same virus were tested for human homology. Specifically, the HKHD40 and HKNPC60 variants were often considered together as "EBV."

Methods Used for Chromosome Comparisons of Breakpoints in Breast Cancers in High-Risk Women Versus Breakpoints in Sporadic Breast Cancer

The NCBI Genome Decoration page provided chromosome annotation software [46].

Identifying Genes Around the Most Frequent EBV-Binding Site Locations and Tethering Sites

EBV nuclear antigen 1 (EBNA1)–binding location genome coordinates [47,48] were used to tabulate genes within or near anchoring sites where EBV docks on human DNA. Breaks in breast cancers were compared to the gene positions around their EBNA1-binding sites. The Palindrome Site Finder from NovoPro and the EMBOSS palindrome program were used to identify palindromic DNA sequences.

Comparisons for Similarities Among Human Herpesviruses

EBV variants HKHD40 and HKNPC60 were compared to human herpesviruses in BLASTn by entering the terms "human gamma herpesvirus 4," "herpesviridae," and "herpesvirales." Values with ≥2000 bp in common were selected. The EBV reference sequence was also tested against the following proven cancer viruses: human herpesvirus 8 (also called Kaposi sarcoma virus), herpes simplex virus 1, and human cytomegalovirus.

Locating Piwi-Interacting RNA Sequences as Evidence of Past EBV Infection

Piwi-interacting RNA (piRNA) locations were retrieved from the piRNA bank [49,50]. To compare the positions of piRNAs in virus homology versus genome position graphs, the midpoints of piRNA sequences were assigned arbitrary homology values. Positions of differentially methylated regions near breast cancer breakpoints on chromosome 6 [51] were compared to breakpoint positions for 70 NPCs based on published data analyses [11].

Viral Sequences in Human Genomes as Hypermutation and Rearrangement Sites in Breast Cancers

A graph of viral sequences in humans against chromothripsis breaks in breast cancers was so complex that it resisted interpretation, so only the 5 viral sequences nearest the chromothripsis breaks were used. The viral sequences nearest high-confidence chromothripsis breaks were determined in 5 iterations as genome coordinates where *XLOOKUP* values gave the minimum distances. Distances between all virus homology start points were then compared to all chromothripsis breakpoints.

Methods of Data Analyses and Statistical Software

DNA flanking sequences at breakpoints were downloaded primarily from the GRCh38/hg38 version of the University of California, Santa Cruz Genome Browser as FASTA files and copied directly into BLAST. Results were checked against breakpoints in 101 triple-negative breast cancers from a population-based study [32]. The University of California, Santa Cruz Genome Browser's *Liftover* function interconverted different versions of genome coordinates into GRCh38/hg38 coordinates.

Statistics

Excel, SPSS (IBM Corp), StatsDirect, Visual Basic (Microsoft), and Python (Python Software Foundation) scripts were used for data analysis. Mann-Whitney *U* tests compared overall breakpoint distributions [52] and tested the hypothesis that breakpoint distributions were identical or at least roughly the same. The Mann-Whitney *U* test was chosen because the comparisons involved unequal numbers of breakpoints, and each observation was likely independent. *P* values >.05 were taken to indicate that identical distributions could not be excluded. Tests for normality included kurtosis and skewness values and evaluation by Shapiro-Francia and Shapiro-Wilk methods [53] (Multimedia Appendix 2). The Fisher exact test compared breakpoints in breast cancers to those in known viral cancers. The unpaired 2-tailed Student *t* test was used to compare the means of numbers of breast cancers with severe

versus nil lymphocyte infiltrates, assuming the data approximated normality and that there were no extreme outliers. Both of these tests require independence and random sampling. All these test results are only approximate because they depend on underlying assumptions.

Fragile Site Sequence Data

Positions of common fragile sites were retrieved from a database [54] and original publications [55].

Ethical Considerations

This study presents analyses of publicly available data without recruiting additional human or animal subjects. Because this study is a secondary analysis, it is exempt from institutional review board and ethics approval. The data are in the public domain and are available for independent research and analysis [56]. It is not necessary to obtain permission to reuse public data. The original informed consent allows secondary analysis without additional permission.

Results

Breakpoints in Breast Cancers From High-Risk Backgrounds Clustered Around Breakpoints in NPC, an EBV-Mediated Cancer

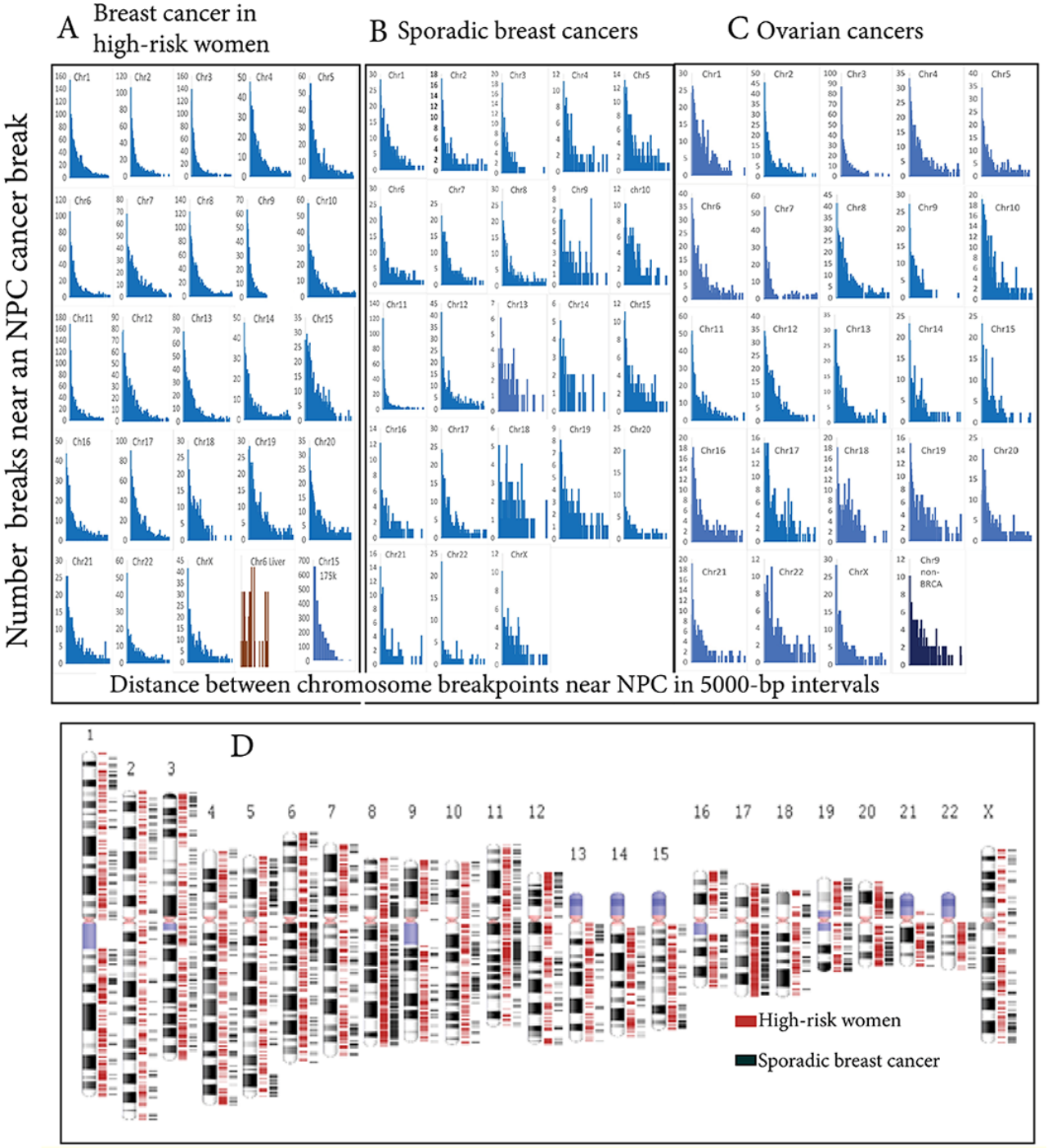
EBV-mediated cancers such as NPC have defects in DNA repair and inflammatory pathways, resembling hereditary breast and ovarian cancers. To further characterize this resemblance, breakpoints in 70 NPC genomes were compared to breakpoints in 139 breast cancer genomes from high-risk women (*BRCA1*/*BRCA2* mutation, familial concentration, or young age).

The distances from all breast cancer breakpoints to the nearest NPC breakpoints across the entire length of chromosome 1 produced results with so many points that they were difficult to interpret (Multimedia Appendix 3). Different laboratories collected these breakpoint data over many years. To allow for some variations, the data were grouped into 5000-bp increments (2×10^{-5} relative error). As shown in Figure 1 and Multimedia Appendix 2, breast cancer breakpoints were most often clustered within 5000 bp of NPC breakpoints, but many breakpoints agreed much more closely. A total of 20 breast cancer breakpoints on chromosome 1 were within 500 bp of an NPC breakpoint, and several chromosomes had breast cancer and NPC breakpoints in essentially the same positions. As represented by Mann-Whitney *U* test results (Multimedia Appendix 2), breast cancer and NPC breakpoint distributions were statistically the same ($P>.05$) for chromosomes 6, 7, 10, 13, 14, 15, 22, and X, but different on chromosome 1 and other chromosomes ($P<.05$).

In contrast, liver cancer breakpoints at hepatitis B virus integration sites [57] differed from those in breast cancer or NPC (Figure 1). No breaks in 114 liver cancers on chromosome 1 were within 5000 bp of breaks in any NPC; only one break on chromosome 6 in 61 liver cancers fit this window. According to a meta-analysis, the chance that breakpoints on chromosomes 1, 2, 6, and 8 were not within 5000 bp in liver cancer versus NPC was 4.4 (95% CI 1.9 - 10). NPC and liver cancer did not have the same breakpoint distributions ($P<.001$).

The above results revealed that breast cancer breakpoints in high-risk women were clustered near those in the EBV-associated cancer, NPC, on every chromosome. The next step was to decide whether these similarities depended on mutations in the breast cancer susceptibility genes, *BRCA1* or *BRCA2*, by comparisons to sporadic breast cancers. The sporadic breast cancer group comprised 74 women, aged ≥70 years, with normal *BRCA* genes and no other known inherited, cancer-associated mutations [31]. Like breakpoints from high-risk women, many sporadic breast cancer breakpoints clustered around those in NPC (Figure 1). Breakpoints in these sporadic breast cancers clustered at chromosomal locations similar to breast cancers from high-risk women, although the frequencies and distributions sometimes differed significantly. The patients with sporadic breast cancer were older than the high-risk women, arguing against age as responsible for similarity to NPC breakpoints.

Figure 1. (A) Breakpoints in 139 breast cancers from high-risk women (*BRCA* mutation, familial concentration, or early onset) clustered around breakpoints in 70 NPCs. The data were grouped in 5000-bp increments to allow for methodological and laboratory differences. An unrelated set of hepatocellular data associated with hepatitis B insertions did not show a similar relationship to NPC. Breast cancer and NPC breakpoint distributions could not be confidently distinguished (P>.05) for chromosomes 6, 7, 10, 13, 14, 15, 22, and X (Multimedia Appendix 2). Many breakpoints were virtually the same on some chromosomes. The panel at the lower right shows how the selection of a larger bin size of 175,000 bp (the approximate length of EBV) affects the distributions of breakpoints. (B) Like the breast cancers from high-risk women, breakpoints in 74 sporadic breast cancers clustered around the breakpoints found in 70 NPCs. Breast cancer breakpoints within 5000 bp of an NPC breakpoint were the largest single category on most chromosomes. (C) Breakpoints in 25 mixed adenosquamous ovarian cancers also clustered around breakpoints in the 70 NPCs. The data show both *BRCA*-associated and nonassociated ovarian cancers. The panel in the lower right corner represents chromosome-9 data after removing all *BRCA*-associated ovarian cancers. The sporadic cancers show the same results as the complete set but with less data. (D) Many breakpoints in sporadic breast cancers clustered at chromosomal locations similar to those from high-risk women. Interchromosome translocation break positions in 74 mutation-associated, familial, or early-onset female breast cancers (red) versus 74 likely sporadic female breast cancers (black) are shown. bp: base pairs; Chr: chromosome; EBV: Epstein-Barr virus; NPC: nasopharyngeal cancer.



Viral Homologies Around Breakpoints in Mixed Adenosquamous Ovarian Carcinoma Also Clustered Around Breakpoints in EBV-Mediated Cancer

Ovarian cancer data enabled an additional test for EBV involvement in breast cancer because, like breast cancer, *BRCA1* or *BRCA2* mutations can also predispose patients to ovarian cancer [58]. Chromosome breakpoints in 25 mixed adenosquamous ovarian cancers were compared to breakpoints in NPCs. The results depicted in Figure 1 emulated breast cancer comparisons. Nearly half (12/25, 48%) the ovarian cancer cases had likely hereditary *BRCA* mutations. The remaining sporadic ovarian cancers gave the same results as the complete set but with less data. As in breast cancer, ovarian cancer breakpoint distributions clustered around NPC breakpoints, even without a hereditary *BRCA1* or *BRCA2* gene mutation driver.

Breaks in Lymphomas Associated With EBV Infection Also Matched Breast Cancer and NPC

EBV drives lymphomas as well as NPCs. Based on epidemiologic research results, FA-BRCA pathways protect against lymphomas [59,60]. If EBV is genuinely associated with breast cancer breakpoints, then breakpoint positions in EBV-mediated lymphomas should also resemble those of breast and ovarian cancers. Because *MYC* gene rearrangements are characteristic of EBV-associated lymphomas, the first test of this idea was to survey virus-like sequences surrounding the *MYC* gene locus on the human reference genome. Figure 2 shows that *MYC* resides in a literal forest of retrovirus sequences (eg, human immunodeficiency virus type 1 [HIV1], feline leukemia virus, porcine endogenous retrovirus, and human endogenous retrovirus [HERV]) interspersed with EBV-like sequences.

The concentration of virus sequences around *MYC* on chromosome 8 prompted the addition of the EBV-associated lymphoma DLBCL to breakpoint comparisons. As shown in Figure 2, the results revealed that hundreds of breast cancer and NPC breakpoints congregated around breakpoint positions in 88 DLBCLs [22]. This agreement was consistent with other similarities between breast cancers and these EBV-associated cancers, including deficits in FA-BRCA pathway–mediated DNA repair by homologous recombination [61] and NF-κB activation [11,62-64].

EBV is also a proven driver of at least one subset of BLs, typically those with *MYC* translocations. BL subsets can have mutations that impair homologous recombination [65], so results in Figure 2 revealed many breast cancer breakpoint positions near corresponding BL breakpoints. An older dataset from BLs [21] had translocation breakpoints in the virus sequence–rich area near the *MYC* locus, agreeing with about 140 breast cancer breakpoints. Four different NPC breakpoints produced over 100 matches to BL translocation breakpoints, beginning at 8250 bp apart. An unpaired, 2-tailed *t* test did not support a statistically significant difference between BL and NPC breakpoints in this area (P=.69).

Further tests were conducted to determine whether the functions of genes near clustered breakpoints supported a relationship between breast cancers and EBV-related cancers (GC [41], BL, and NPC). As illustrated in Figure 3, breast cancer breakpoints on chromosomes 6, 8, 11, and 17 aggregated near positions where breakpoints occurred in EBV-associated cancers. Many aggregated breakpoints were in the same areas as genes that control inflammation, antiviral defenses, apoptosis, intermediate filaments, epigenetic and chromatin regulation, estrogen receptor activity, mitotic structures, and mitotic controls (Table S1 in Multimedia Appendix 4). Breast cancer breakpoints that clustered around EBV-associated cancer breakpoints were especially numerous on chromosome 17. One of these clusters marked in Figure 3 included the HER2 amplicon and the topoisomerase 2a gene, with *BRCA1* and *SMARCE1* genes nearby. *SMARCE1* encodes a part of a chromatin regulation complex. Chromosome 17 breakpoints near *CNTROB* and *CTC1* genes connect EBV to centriole and telomere malfunctions during mitosis (Table S1 in Multimedia Appendix 4). Rearrangements near breakpoints may cause over- or underexpression of nearby genes (Table S2 in Multimedia Appendix 4). Many additional correlations were also likely revealed in Figure 3 but were not investigated further.

Results in this section show that breast cancer breakpoints clustered around breakpoints in additional EBV-associated cancers, where they affect critical functions needed to prevent breast cancer. Once these functions are compromised, cancer can occur without the continuing presence of EBV.

Figure 2. (A) Human DNA around the *MYC* locus on chromosome 8 was filled with virus-like sequences. *CASC11* is an RNA gene that several cancers overexpress. Breast cancer and lymphoma breakpoints were dispersed throughout the *MYC* region and beyond, but NPC breakpoints were less common. (B) On chromosome 8, hundreds of breakpoints in breast cancers and NPCs clustered around breakpoints in data from 88 patients with DLBCL who were likely EBV positive. This agreement highlights multiple similarities among these cancers. (C) EBV drives a subset of BLs, typically with *MYC* translocations and impaired homologous recombination. Based on *MYC* fusion sequences in BL, breast cancer breakpoints on chromosome 8 also clustered around BL breakpoints. BLs from an older dataset [21] had translocation breakpoints in the virus-rich area near the *MYC* locus, agreeing with ≥140 breast cancer breakpoints. *MYC* locus translocations had not been reported in NPCs, but NPC breakpoints still clustered around BL fusion breakpoints, although at greater distances. Four different NPC breakpoints produced over 100 matches to BL translocation breakpoints beginning at about 8250 bp apart. An unpaired, 2-tailed *t* test did not support a statistically significant difference between BL and NPC breakpoints in this area (*P*=.69). BL: Burkitt lymphoma; bp: base pairs; DLBCL: diffuse large B-cell lymphoma; EBV: Epstein-Barr virus; FeLV: feline leukemia virus; HERV: human endogenous retrovirus; HERVK: human endogenous retrovirus K; HIV1: human immunodeficiency virus type 1; HPV18: human papillomavirus 18; HRV: human retrovirus; NPC: nasopharyngeal cancer; PERV: porcine endogenous retrovirus; Stealth: stealth virus 1.

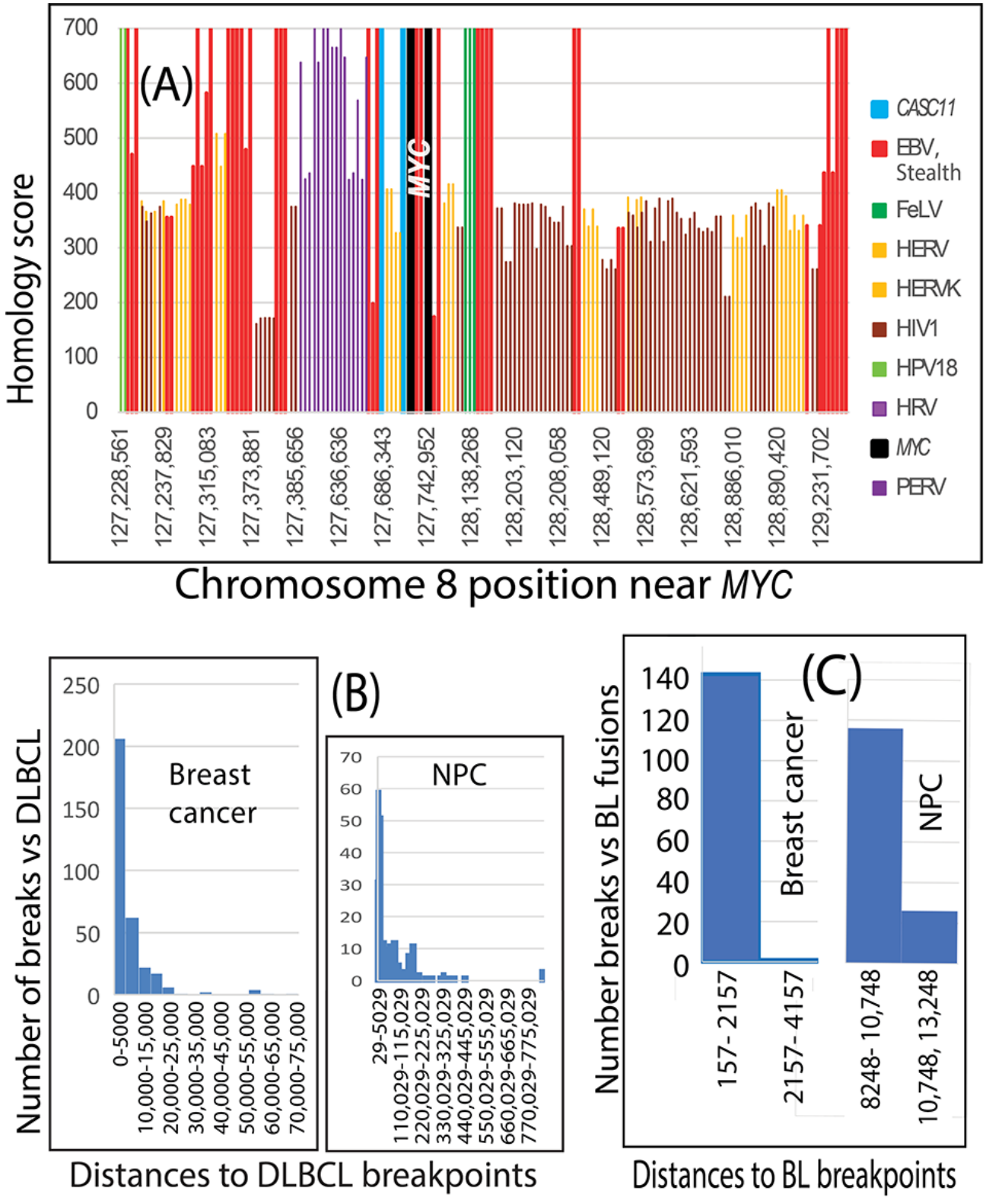
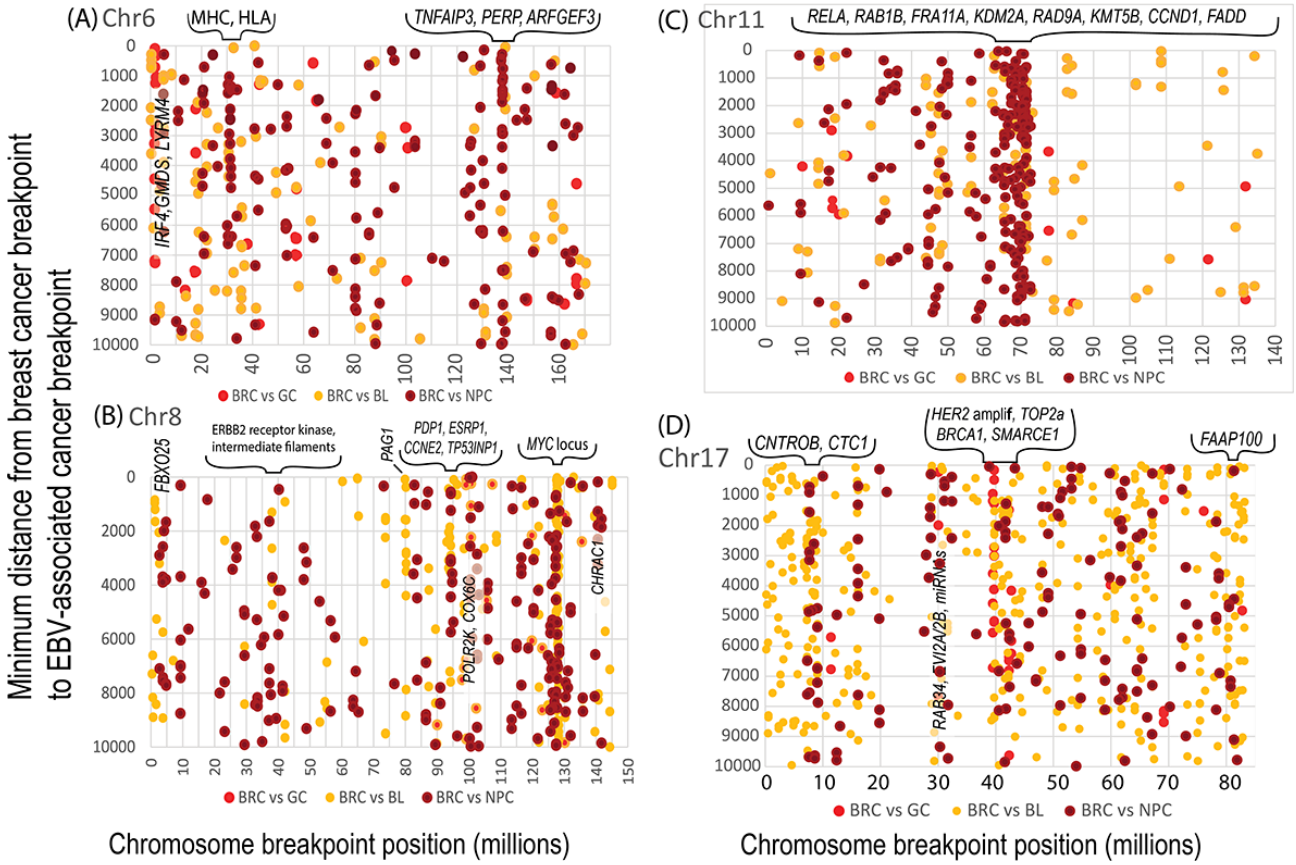


Figure 3. Breakpoints in breast cancers clustered around breakpoints in EBV-positive cancers in 3 different tissues. The EBV-positive cancers comprised 34 GCs, 90 BLs, and 70 NPCs. The clustering of breast cancer breakpoints and EBV-related cancer breakpoints was pronounced on chromosomes (A) 6, (B) 8, (C) 11, and (D) 17. Selected genes around some of the clustered breaks are indicated. Functions of the genes can have profound effects on the human genome and are summarized in Table S1 in Multimedia Appendix 4. BL: Burkitt lymphoma; BRC: breast cancer; Chr: chromosome; EBV: Epstein-Barr virus; GC: gastric cancer; HLA: human leukocyte antigen; MHC: major histocompatibility complex; NPC: nasopharyngeal cancer.



Genes at the Most Frequent EBV-Tethering Sites Clustered Around Breast Cancer Breakpoints

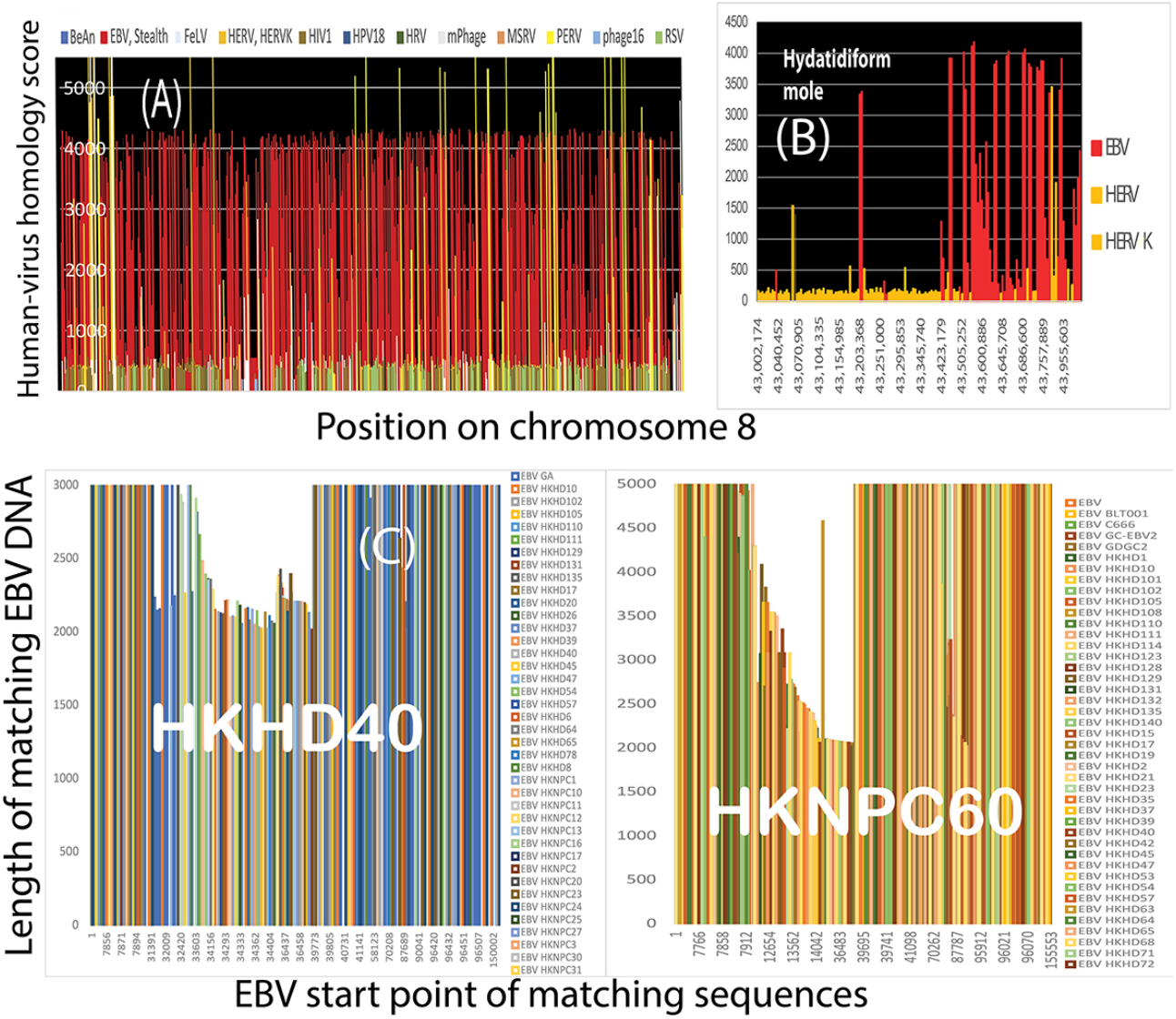
In preceding sections, breast and ovarian cancer breakpoints were found to distribute most frequently near characteristic sets of breakpoints associated with EBV-related cancers. The virus first attaches its EBNA1 protein to human DNA in the nucleus. Then, circular EBV episomes dock to this attached EBNA1 anchor. To test whether the initial EBNA1 attachment sites were related to breast cancer chromosome breakpoints, breast cancer breakpoints were compared to genes near EBV-docking sites. EBV-positive BL cells providing the data had up to 1569 EBV-docking sites on all chromosomes identified by 4C-chromatin capture experiments [47]. As shown in Figure 4A, the largest numbers of breast cancer breakpoints on most chromosomes clustered around the genes [47] nearest to genes at EBV-docking sites. In support of these comparisons, graphical estimation of virus-tethering sites on chromosome 2 from chromatin capture data for these EBV-positive cells also agreed with breast cancer breakpoints (Figure 4A). In an unrelated study [48], EBV-docking sites on chromosome 11 near known EBV anchor sites at the *FAM-D* and *FAM-B* genes were found near groups of breast cancer breakpoints, but imperfect palindrome sequences [66] were more distant (Figure 4B). This finding independently supports the idea that EBV-docking sites are near breast cancer breakpoints. Results in this section raise the possibility that EBV directly contributes to breast cancer chromosome breakpoints and fragmentation.

Figure 4. Relationships of EBV-docking sites to breast cancer breakpoints. (A) Breast cancer breakpoints clustered around the top 10% most frequently found genes near EBV-tethering sites in BL cells. Some of the best information on EBV-docking sites comes from 4C-chromatin capture experiments in EBV-positive BL cells [47]. The largest number of breast cancer breakpoints on most chromosomes clustered around the genes nearest EBV-tethering sites. BL cells providing the data had up to 1569 EBV-docking sites distributed over all chromosomes [47]. EBV-docking sites on chromosome 11 near the *LUZP2* and *FAT3* genes in BL cells were millions of bp from the 18-bp imperfect palindrome interval. Graphical estimation of virus-tethering sites on chromosome 2 (green) from these EBV-positive cells also agreed with breast cancer breakpoints. (B) Independent evidence relating breast cancer chromosome breakpoints to EBV-docking sites. Maximum homology to human DNA for all viruses (y-axis) is plotted around known EBV genome anchor sites on chromosome 11 near the *FAM55D* and *FAM55B* gene coordinates. A posited imperfect palindrome sequence [66] as an EBV-docking site was more distant from the *FAM55* genes. BL: Burkitt lymphoma; bp: base pairs; Chr: chromosome; chrom: chromatin; EBV: Epstein-Barr virus; HERV: human endogenous retrovirus; HERVK; human endogenous retrovirus K; HIV1: human immunodeficiency virus type 1; HTLV1: human T-cell lymphotropic virus type 1; MSRV: multiple sclerosis retrovirus; RSV: respiratory syncytial virus.

Breakpoints Occurred Near Human Sequences That Resemble Viruses in All Breast Cancers Tested

To further test whether EBV itself has some role in breaking chromosomes or altering their structures, human chromosomes were compared to all known viruses. As shown in Figure 5A, the results showed that nearly every breast cancer likely had undergone breakages near EBV-like sequences. Chromosome 8 alone had 59,566 significant (>200) viral homology scores. Based on data from 128 patients with breast cancers and 43,491 unique breakpoints, breakpoints in 123 (96.1%) out of 128 breast cancers were within 10,000 bp of a virus sequence. In 106 patients, the virus was an EBV tumor variant (HKHD40 or HKNPC60) with 3086 matching human sequences. According to the Fisher exact test, chromosome 8 breakpoints and EBV variant sequence matches were not independent (P<.001).

Figure 5. (A) All viral homologies on the entire lengths of chromosome 8 (a total of 145,138,636 bp) are shown in 200k-bp increments. Maximum homology scores over 4000 for human DNA versus herpes viral DNA were abundant. The 4000 score corresponds to 97% human-virus identity over nearly 2500 bp, with *E* ("expect") values (essentially *P* values) effectively equal to 0. The EBV tumor variants, HKNPC60 and HKHD40, were nearly identical to human breast cancer DNA at many positions throughout chromosome 8. (B) It is unlikely that homologies to EBV sequences occurred because the human reference genome was contaminated with EBV episomes. Homozygous hydatidiform mole cells that had lost the paternal chromosomes after fertilization still had strong homology to EBV sequences, such as HKHD40 and HKNPC60 variants. (C) EBV variants HKHD40 and HKNPC60 are typical of hundreds of other EBV variants. Hundreds of human gamma herpesvirus 4 variants are almost identical to HKHD40 and HKNPC60 over at least 2000 bp. The matching sets of viruses included many high-risk herpesvirus isolates from NPCs [67]. BeAn: BeAn 58058 virus; bp: base pairs; EBV: Epstein-Barr virus; FeLV: feline leukemia virus; HERV: human endogenous retrovirus; HERVK: human endogenous retrovirus K; HIV1: human immunodeficiency virus type 1; HPV18: human papillomavirus 18; HRV: human retrovirus; mPhage: mycolicibacterium phage J1; MSRV: multiple sclerosis retrovirus; NPC: nasopharyngeal cancer; PERV: porcine endogenous retrovirus; RSV: respiratory syncytial virus; Stealth: stealth virus 1.



Many areas on other chromosomes also had 97% human-virus identity over nearly 2500 bp. It is implausible that this much similarity comes from EBV DNA being carried over into the human reference genome. Viral homology occurred with only a small, select portion of viral DNA [68]. Viral homologies were determined for a human genome in a homozygous karyotype, haploid cell line (46,XX) hydatidiform mole derived only from the paternal chromosomes in an X-bearing sperm cell after fertilization [69]. Results still showed extensive homology between the mole and EBV variants HKHD40 and HKNPC60 (Figure 5B).

HKHD40 and HKNPC60 variant sequences kept appearing in comparisons to human sequences, so these variants were tested against other herpesviruses to determine whether they were unusual. Hundreds of human gamma herpesvirus 4 variants were almost identical to HKHD40 and HKNPC60 over at least 2000 bp (Figure 5C). The matching sets of viruses included many high-risk herpesvirus isolates from NPCs [67]. Based on this information, HKHD40 and HKNPC60 strongly resembled other herpesvirus isolates, including many that confer high risks for NPC [10]. These results show that humans have interacted extensively with EBV; the results are not due to EBV impurities in the human reference genome, and the human genome has had close relationships with oncogenic EBV forms.

Evidence of Past EBV Infection

The evidence thus far supports a central hypothesis that EBV disables tumor suppressor mechanisms in breast cancer and can then disappear. This absence of viral particles is a significant experimental obstacle to testing this hypothesis. Unlike retroviruses, EBV and its variants do not have integrase enzymes, so EBV has no conventional way to insert itself into the human genome. EBV rarely integrates, with only one or two copies in BL cell lines [70].

BLAST analysis found about 65,000 areas of strong homology ($E<1\times10^{-10}$) between the human reference genome and EBV. Because 65,000 is far more than realistic EBV integration events, it suggested the possibility that some EBV sequences were fragments created by a human version of the bacterial CRISPR (clustered regularly interspaced short palindromic repeats) system. As shown previously in Figure 3, breast cancers have breakpoints that cluster around breakpoints in EBV-associated cancers and involve MHC genes.

MHC genes are encoded on chromosome 6p21.3 in a region that becomes a candidate for such a human CRISPR version. Variants of human leukocyte antigens (HLAs) in the MHC are strong risk factors for NPC infections [71] because HLAs are required to break down and display fragments of some antigens to the immune system. A total of 13 breast cancers listed on the COSMIC website had a deletion near this HLA region. About 23% of breast cancers had mutations directly affecting HLA class I or II genes. Many more breast cancers had indirect connections because they had damage to multiple genes that interact with HLAs or were otherwise essential for immunity. The MHC region also holds *NFKB1L1*, a negative regulator of the NPC overexpressed gene hallmark, NF-κB. The 139 breast cancers from high-risk women had 284 breakpoints at chromosome 6p21.3. Breakpoints in the 70 NPC cancers also clustered there, with 40 breakpoints within the 27,865,296-34,017,013 segment on chromosome 6. Variability in the inactivation of MHC genes reflects the extreme diversity of this region.

In general, the bacterial CRISPR/Cas system loosely resembles the human piRNA system, so the distribution of piRNAs was graphed. As shown in Figure 6A, hundreds of piRNA sequences cluster near the MHC region (at ~29.7 - 33.3 megabases). The piRNA system is known to inactivate virus-derived transposons (related to HERVs) by methylating or cleaving them. The distribution of piRNA fragments was then compared to the distribution of viral DNA fragments in the MHC region of chromosome 6. Figure 6B-F reveals striking similarities in how remnants of exogenous and endogenous viruses distribute relative to piRNAs. Remnants of both virus types were homologous to the same human sequence, and both types were interspaced between piRNA sequences, sometimes right next to each other. Most of these sandwiches were at a regular interval or a multiple of a regular interval.

This interspaced arrangement looked so much like CRISPR that it raised the question of whether piRNA defense mechanisms have inactivated some EBV variants in addition to their canonical role with endogenous viruses. Long stretches of endogenous transposon-like DNA sequences routinely matched exogenous viruses. As shown in Figure 6C and E, the same human DNA interval had homology both to endogenous transposons (HERV) and exogenous viral sequences (EBV variants, stealth virus 1, chikungunya virus, BeAn 58058 virus, human papillomavirus [HPV] 16, HIV1, and HERV). This result shows that the piRNA system can store the same piece of DNA to protect DNA against these different viruses.

Chromosome 6p21.3 also contains an EBV infection marker [72]. The marker was examined in 1538 breast cancers using existing methylation data [34]. As indicated in Figure 7, promoter methylation differed significantly from normal controls in the segment shown (30,523,984 - 33,216,811 on chromosome 6). Hypermethylation occurred on *STK19*, a MHC class III gene for RNA surveillance [73,74]. Hypermethylation also occurred on a gene for preventing tumors (*TNFB*) [75] and a gene for responding to antigen-antibody complexes (*C2*). Polymorphisms in *HLA-DMB* antigen and *SAPCD1*, another class III MHC gene [76], at chromosome 6p21.3 had links to Kaposi sarcoma [77]. Human herpesvirus 8 (Kaposi sarcoma virus) is a Kaposi sarcoma driver and is closely related to EBV.

These results reveal that EBV has been attacking human DNA during evolution. There is a piRNA defense mechanism for human DNA near critical immune system genes, but both EBV-associated cancers and breast cancers inactivate some of the genes that guard piRNA defenses. The histocompatibility antigen gene region of chromosome 6 can be extensively fragmented in EBV-associated and breast cancers. MHC genes have the largest number of polymorphic forms in the human genome. This variation creates differences in viral susceptibility and inactivation. Even though most people are infected, not everyone will get an EBV-related disease or cancer.

Figure 6. The human genome organizes piRNA sequences into clusters near the MHC region of chromosome 6 (6p21.3 at ~29.7-33.3 megabases), with hundreds of piRNAs nearby. (A) The levels of various piRNAs varied by more than 1000-fold, but the most abundant piRNAs were the only ones present in every cell. These abundant sequences drive the inactivation of foreign DNA. Rare piRNAs do not function in every cell but can potentially adapt to new genome invaders. (B-F) Arbitrarily selected areas of the chromosome region where piRNAs are most abundant. piRNAs were assigned sufficient homology scores to mark their positions relative to positions with homology to viruses. (C and E) Remnants of both exogenous and endogenous virus types were homologous to the same human sequence, and both types were sandwiched between piRNA sequences, sometimes right next to each other. Most sandwiches were at a regular interval or a multiple of a regular interval. The same human DNA interval has homology to endogenous transposons (HERV) and exogenous viral sequences (ChikV, HIV1, Stealth, BeAn, and HPV16). The piRNA system can store the same piece of DNA to protect DNA against these different viruses. BeAn: BeAn 58058 virus; ChikV: chikungunya virus; Chr: chromosome; EBV: Epstein-Barr virus; HERV: human endogenous retrovirus; HERVK; human endogenous retrovirus K; HIV1: human immunodeficiency virus type 1; HPV16: human papillomavirus 16; MHC: major histocompatibility complex; PERV: porcine endogenous retrovirus; piRNA: Piwi-interacting RNA; Stealth: stealth virus 1.

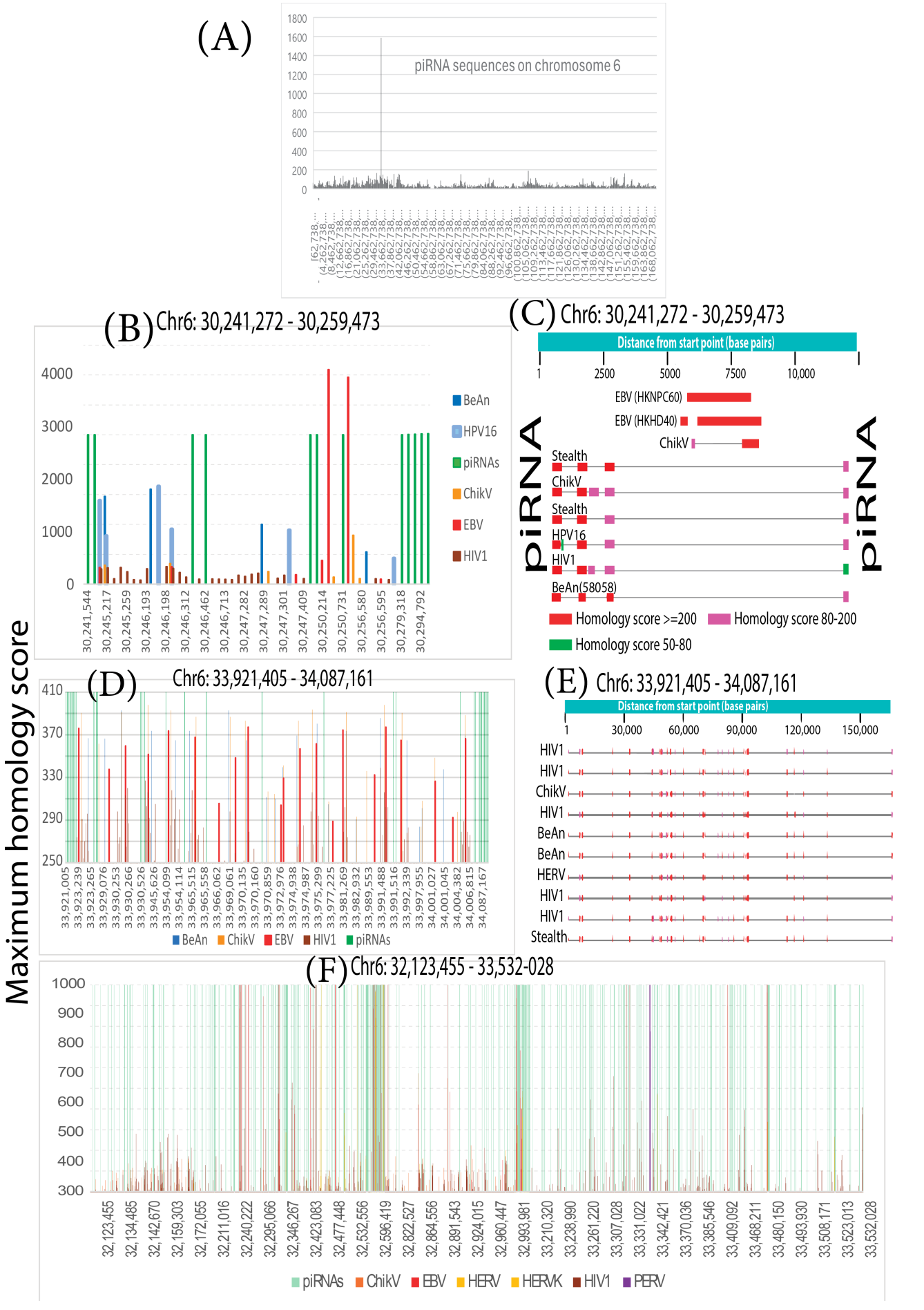
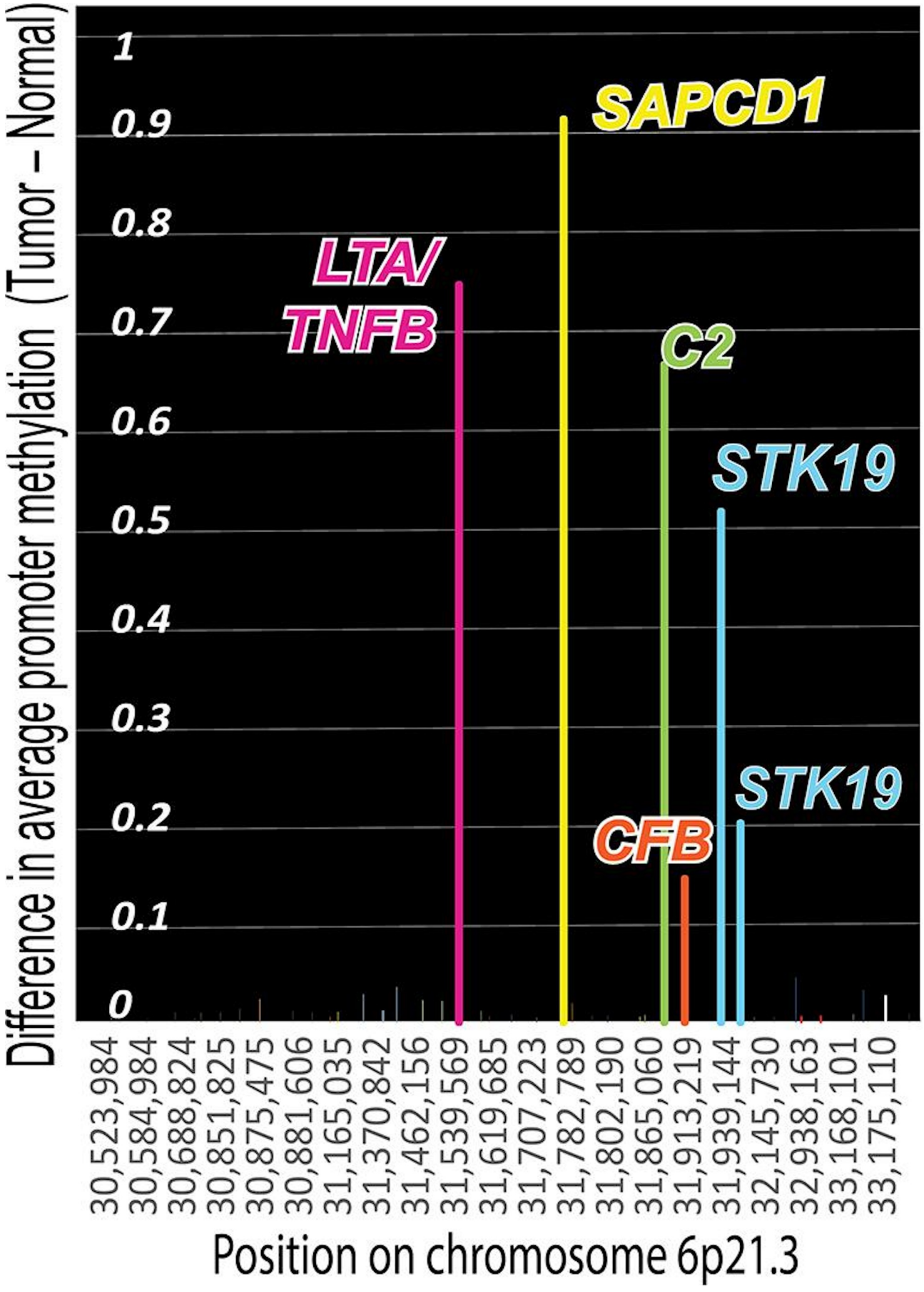


Figure 7. Chromosome 6p21.3 contains an EBV infection signature [72]. Using existing methylation data [34], the marker was examined in 1538 breast cancers. Promoter methylation in this marker region differed significantly from normal controls. Hypermethylation occurs on *STK19*, an MHC class III region gene [73] for RNA surveillance [74]. Hypermethylation also inhibited *LTA/TNFB*, a gene for preventing tumors [75], and *C2*, which encodes antigen-antibody complex responses. Polymorphisms in *HLA-DMB* antigen and *SAPCD1*, another class III MHC gene [76], at chromosome 6p21.3 have links to Kaposi sarcoma [77]. HHV8 is a Kaposi sarcoma virus closely related to EBV. EBV: Epstein-Barr virus; HHV8: human herpesvirus 8; HLA: human leukocyte antigen; MHC: major histocompatibility complex.

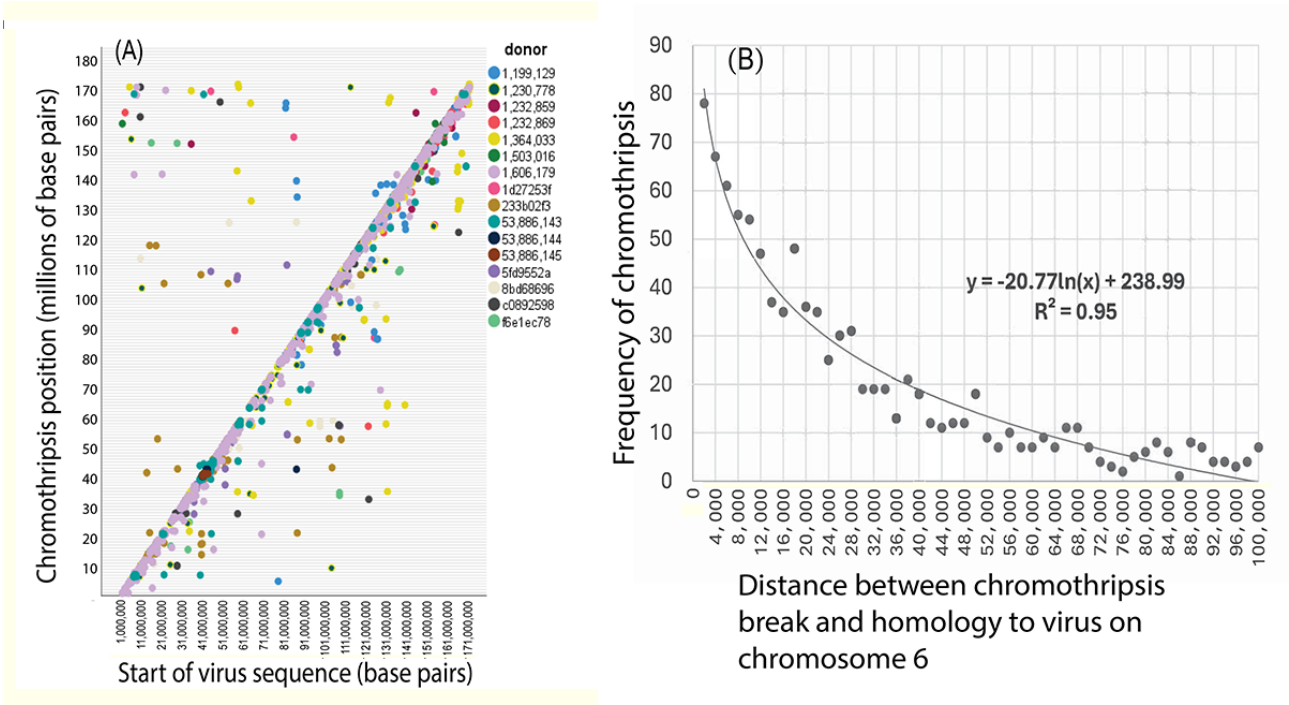


Viral Sequences in Human Genomes as Hypermutation and Rearrangement Sites in Breast Cancers

The next question was whether EBV or other virus-like sequences in the human genome cause multiple rearrangements and clustered hypermutations (chromothripsis). As shown in Figure 8A, many positions on chromosome 6 where chromothripsis occurs [35] congregated around virus sequence start positions. A total of 1090 genome coordinates described chromothripsis fragments with copy number ≥3. These coordinates were unlikely to be random since they did not follow a normal distribution (P<.001). By simple linear regression analysis (R^2=0.93), many viral sequence coordinates strongly correlated with chromothripsis positions. Figure 8B shows that as you move further away from a chromothripsis breakpoint, the frequency of breast cancer homology (score>500) to viruses decreases. This result implicates viral sequences as preferred sites where breast cancer chromosomes begin to fall apart. The equation shown mathematically describes the relationship between chromothripsis frequency and distance from viral sequences, and the constant in the equation suggests a baseline level of breakpoints.

These results suggest that homologous virus sequences at multiple positions could confuse DNA repairs already compromised by EBV in breast cancer and contribute to chromothripsis and clustered rearrangements.

Figure 8. Repetitive copies of virus sequences may confuse compromised DNA repairs and contribute to hypermutation clusters and rearrangements. (A) High-confidence positions where chromosome 6 shatters in 16 breast cancer genomes [35] were plotted against start points of viral sequence homologies. EBV or other viruses then cause groups of rearrangements and hypermutation clusters (chromothripsis). A total of 1090 genome coordinates described fragments with copy number ≥3. These coordinates were unlikely to be completely random since they did not follow a normal distribution (P<.001). Genome coordinates on chromosome 6 matching virus sequences were strongly correlated by simple linear regression analysis (R^2=0.93). (B) As you move further from chromothripsis breakpoints, the frequency of breast cancer homology to viruses decreases, according to the equation shown. The constant in the equation suggests a baseline level of breaks.



EBV and Metastasis

The last question was whether EBV contributes to breast cancer metastasis. According to Yates et al [29], relapsed and metastatic breast cancer tumors keep their tumor-driver gene mutations and continue acquiring new ones. Late mutations in JAK-STAT and SWI-SNF signaling pathways drive established breast cancers into metastasis.

NPC often loses type-1 interferon genes (*IFNA1*, *IFNA2*, *IFNA8*, and *IFNE*) and nearby *MTAP* (32%-34% [11]) by homozygous deletions at chromosome 9p21.3. Interferons initiate canonical JAK-STAT signaling by binding to cell surface receptors that then activate internal Janus kinases (JAKs). The activated JAKs phosphorylate cytoplasmic STAT (signal transducer and activator of transcription) proteins, which travel to the cell nucleus to activate interferon-responsive genes. The percentages of breast cancers on the COSMIC website with mutations in a "JAK" or "STAT" isoform or transcript variant were calculated: 7.8% had a JAK mutation and 36.7% had a STAT mutation. Deletions of interferon genes in NPC also facilitate viral replication and block interferon from activating JAK-STAT signaling. Breast cancers (Multimedia Appendix 2) have 65 breakpoints strictly within this interferon-*MTAP* region (21,579,478 - 20,503,534 on chromosome 9), not counting longer fragments that include the interval. As shown in Figure 9, breast cancer breakpoints align well with EBV-associated cancer breakpoints near the large cluster of interferon genes on chromosome 9.

Figure 9. Damage to JAK-STAT and SWI-SNF signals pushes breast cancer into metastasis [29]. EBV interferes with these signaling pathways to facilitate viral replication. (A) Breakpoints in breast cancers on chromosome 9 facilitated viral replication and blocked sources of JAK-STAT signaling, including a large cluster of interferon genes on chromosome 9. Breast cancers can disable SWI-SNF by targeting *ARID* genes. (B) *ARID1A* was encoded on chromosome 1 near a hot spot where multiple breast cancer breakpoints approximately aligned with breakage points in EBV-associated cancers. Another site at about 150,000,000 bp had a histone-rich region nearby. SWI-SNF affects histones, which also profoundly affects metastasis [78]. The GRCh38 genome version does not include centromere sequences due to technical limitations. *ANXA1*: Annexin A1; BL: Burkitt lymphoma; bp: base pairs; BRC: breast cancer; Chr: chromosome; EBV: Epstein-Barr virus; GC: gastric cancer; NPC: nasopharyngeal cancer; SWI-SNF: switch/sucrose non-fermentable.

Mutations in EBV-associated cancers show that Yates metastasis driver gene damage accompanies EBV infection. SWI-SNF (switch/sucrose non-fermentable) is a complex that repositions nucleosomes and supports genome stability [79]. SWI-SNF addresses obstacles to replication sensed by the FA-BRCA pathway [79,80]. Referring back to Figure 3, clustered breast cancer breakpoints on chromosome 17 around EBV breakpoints affect the SWI-SNF component *SMARCE1*. In addition, breast cancers can disable SWI-SNF by targeting *ARID* genes [29]. *ARID1A* is a COSMIC top-20 most frequently mutated gene in breast cancer. Like breast cancer, NPC has multiple recurrent aberrations in *ARID1A* genes. As shown in Figure 9, *ARID1A* lies near a hot spot where multiple breast cancer breakpoints approximately aligned with breakage points in EBV-associated cancers. The loss of *ARID1A* activates Annexin A1, which aligned closely with a region targeted by EBV-associated cancers on chromosome 9. A chromosome-1 site at about 150 million bp had a nearby histone-rich gene region. Histones are chromatin structures that SWI-SNF dynamically remodels to regulate access to genetic information. Histones can profoundly affect metastasis [78]. Figure 9 also reveals many additional alignments between breakpoints in breast and EBV-associated cancers that were not investigated further.

NPC often inactivates SWI-SNF components *BAP1* and *PBRM1* within a frequently damaged 3p21.3 gene cluster [11] at 52,400,000 - 53,000,000 on chromosome 3. Analyses of breast cancers found 18 breakpoints within this short interval. DLBCL, another EBV-linked cancer, also had recurrent alterations in components of SWI-SNF complexes [81].

The Warburg effect (oxidative glycolysis) [68] favors metastasis. The Warburg effect occurs in NPC because pyruvate dehydrogenase *(PDHB)* genes on chromosome 3p are deleted or rearranged in almost all cases. Similar changes to chromosome 3p were found in breast cancers, which also undergo the Warburg effect [68]. This Warburg metabolic switch favors metastasis because it mitigates oxidative stress on cancer cells. Large amounts of lactate accumulate in the absence of *PDHB* to acidify the tumor microenvironment and interfere with the destruction of metastatic cells [82].

This section's results show that EBV may push breast cancer into metastasis by interfering with JAK-STAT and SWI-SNF signaling pathways to facilitate viral replication while making the microenvironment more favorable to tumor growth.

Alternative Explanations for Breast Cancer Breakpoints That Do Not Involve EBV Variants

Subgroups

To determine whether breakpoint similarities in viral and breast cancers depended on specific subgroups, relationships to NPC were compared in triple-negative and HER2-positive breast cancers (20 and 22 patients, respectively). Triple-negative breast cancers are likely to be *BRCA1* mutation positive [83], while HER2 amplification is uncommon in *BRCA1* and *BRCA2* mutation carriers [84]. Although subgroup differences are noticeable, results still show that both subgroups had breakpoints on all chromosomes related to NPC (Figure S1A in Multimedia Appendix 5).

Tumor-Infiltrating Lymphocytes

Tumor-infiltrating lymphocytes (TILs) are biomarkers for predicting breast cancer prognosis [85,86]. To test whether TILs cause chromosome breaks, breakpoint numbers in 16 breast cancers with severe lymphocyte infiltration were compared to 17 breast cancers with nil lymphocyte infiltration. The 2-tailed Student *t* test could not reject the null hypothesis that the numbers of breakpoints were statistically identical (P=.70; Figure S1B in Multimedia Appendix 5). This result does not rule out differences in prognosis due to differences in lymphocyte infiltration.

Retroviruses

Retrovirus contributions to structural variations were estimated using data from cancer in 38 different tissues [87]. Retrotransposons make relatively modest contributions to breast cancer compared to, say, esophageal or oral (gums) cancer (Multimedia Appendix 5). EBV can transactivate endogenous retroviruses [11,87,88]. DNA near some breast cancer breakpoints resembles porcine endogenous retrovirus, HERV, and HIV1 (eg, Figure 5). The human genome also contains DNA matching the retrovirus mouse mammary tumor virus [89,90] at 23 sites that give BLAST homology scores>200. HPV variants are DNA viruses that are also implicated in breast cancer. HPVs were not assessed further, but they occasionally matched DNA near breast cancer breakpoints.

Common Fragile Sites

Common fragile sites are site-specific breaks seen on metaphase chromosomes after inhibiting DNA synthesis via DNA polymerase inhibitors. Some common fragile sites [54] aligned with breast cancer breaks on chromosome 1, but breakpoints on most other chromosomes were incompatible. Chromosomes 8, 9, 11 - 15, 17-19, 21, and 22 do not have common fragile sites but still have many breast cancer breaks [91]. However, the human genome has over 13 million palindromes that are ≤40 bp [92]. The generation of rare fragile sites by palindromes or their attraction to EBV cannot be excluded.

Imperfect Palindrome Repeats

An alternative explanation for EBV-related carcinogenesis involves the docking of EBNA1 virus-tethering protein at imperfect palindromes [93] tandemly repeated on chromosome 11. The docked EBNA1 binds EBV circular episomes, and chromosome 11 breaks initiate malignancy. To test this explanation, existing literature data were first compared to the specific human EBNA1-binding site [48,66,94]. The results (Table S2 in Multimedia Appendix 4) are incompatible with a single host sequence binding EBNA1.

BLAST analysis showed that matches to the imperfect palindrome were likely due to pure chance with E values between 16 and 964 for 4352 matches, from 12 to 18 bp. Chromosome 11 had only 197 of these 4352 matches, and none were near the palindromic region. The prototype DNA palindrome (Table S2 in Multimedia Appendix 4, line 2) produced 7074 matches with E values ranging from 0.25 to 964. Further BLAST analyses of the slightly different docking sequence in EBNA1-DNA crystals (Table S3 in Multimedia Appendix 4, line 1) against other genome assemblies [95]

revealed matches on chromosomes 2, 19, 4, and 12. Various isolates of HIVs had 52 matching sequences.

In 94 BL samples from patients who were EBV positive, breakpoints concentrated within chromosomes 2, 8, 13, 14, and 22 (Figure S1D in Multimedia Appendix 5). Chromosome 14 contained 610 breakpoints (*IgVH* regions), and chromosome 2 (*IgVK* regions) contained 522 breakpoints. EBV hijacks activation-induced cytidine deaminase, a mutagenic enzyme that generates antibody gene variants in response to myriad antigens. In the 94 EBV-positive BL cases, the palindromic locus was nearly 100 million bp away from the principal breakpoint coordinates (Figure S1E in Multimedia Appendix 5). Only 19 (20%) of the 94 patients who were EBV positive [96] had breakpoints anywhere on chromosome 11. The palindromic locus was also not involved in diverse cancers from 8227 patients [97] (Multimedia Appendix 5).

The results in this section show that alternative explanations that invoke subgroups, TILs, retroviruses, or a specific palindromic repeat locus are incompatible with the associations between EBV-associated and breast cancers .

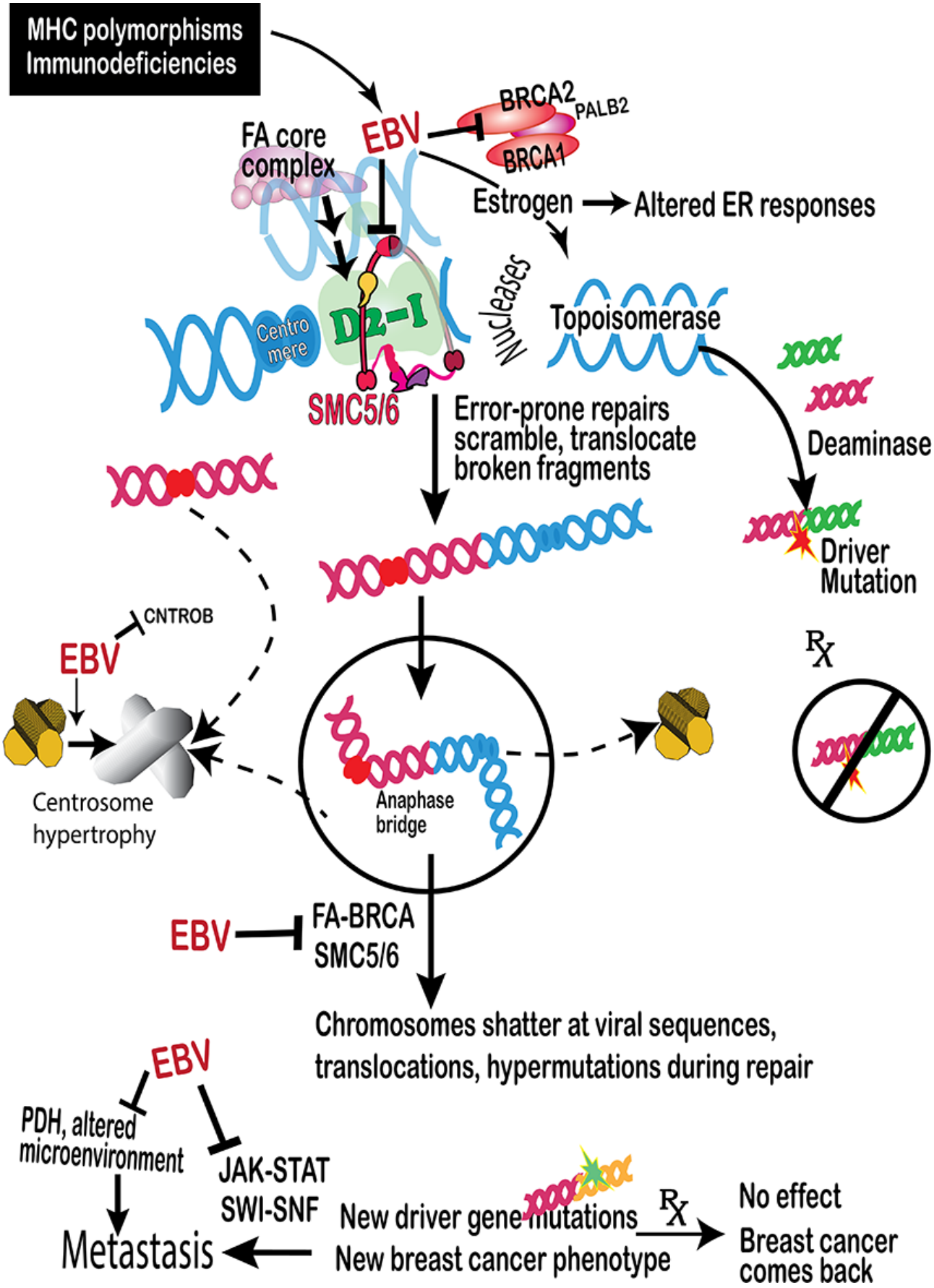
Discussion

Principal Findings

This study finds that EBV contributes to breast cancer by disabling safeguards against tumors. Cancer then occurs because the safeguards remain disabled even if the virus is cleared. Multiple independent analyses identified residual genetic and epigenetic damage in cancer genomes and formed the basis of the model in Figure 10. Breakpoints in breast cancers in high-risk women, sporadic breast cancers, and even ovarian cancers cluster around breakpoints in known EBV-related cancers, including NPC, BL, DLBCL, and GC. Some genes clustered near breakpoints in these diverse EBV-associated cancers are critical to preventing breast cancers. Some breast cancer breakpoints are near genes at EBV-docking sites. Varying numbers of DNA breaks occur within the highly polymorphic forms of MHC region genes on chromosome 6. This damage adds to susceptible polymorphisms and immunodeficiencies to help explain why not everyone develops EBV-related cancers. Near the MHC region on chromosome 6, piRNA sequences are regularly interspaced between viral DNA sequences. The sandwiched arrangements are presumptive evidence of past infection and probably represent a DNA defense mechanism. These defenses fail when chromosome 6 breaks apart near start points of the large number of repetitive viral sequences in the human genome. The viral sequences confuse repairs already damaged by EBV, and bursts of mutation occur where scrambled fragments ligate. EBV disables the most reliable restoration of broken chromosomes back to their native forms, so repairs form structures with multiple centromeres. These structures undergo additional rounds of fragmentation during cell division. The process continually forms new cancer driver mutations and allows cancer to come back after successful therapy (Figure 10). An EBV methylation signature on chromosome 6 was far more abundant in 1538 breast cancers than in normal controls. Finally, EBV facilitates its own replication by damaging JAK-STAT and SWI-SNF signaling pathways, which pushes breast cancer into metastasis, while virus-associated changes on chromosome 3p interfere with the destruction of metastatic cells. Models [8,98] of EBV-infected human mammary cell cultures transplanted into immunosuppressed mice and EBV loss from NPC cells are consistent with these results.

The study herein has current and future clinical implications in addressing cancers and chronic diseases. An early childhood vaccine against EBV may reduce the incidence of breast cancer on a global scale. If this vaccine even approaches the effectiveness of the HPV vaccine for cervical cancer, then the reduction of breast cancer incidence would be substantial. In breast cancer cases where active infection can be demonstrated, immunotherapy or antivirals can be considered. The results also heighten concern about hidden dangers from viral infections. EBV infection leaves behind persistent genome abnormalities (“long EBV”) linked to breast cancer. Not everyone develops an EBV-related cancer even though almost everyone is infected, suggesting risk assessment should include MHC polymorphisms. MHC genes have abundant connections to both EBV infection [99] and breast cancer [100-102].

Figure 10. Model proposed to explain the results. EBV causes serious disease in only some people due to MHC variants and other damage to the immune system. Viral nucleases are one source of chromosome breaks. EBV causes inappropriate expression of estrogen and transcription targets of occupied estrogen receptors. Transcription induced by artificially high estrogen levels then induces topoisomerase-mediated DNA breaks. EBV-mediated deregulation of estrogen production, topoisomerase activity, and deaminase activation then collaborate to cause chromosome breaks and drive translocations [68]. EBV-associated cancers share additional genome deficits with breast cancers, which interfere with restoring the genome from DNA crosslinks and DNA double-strand breaks. If crosslinks and DNA breaks persist during cell division, they also cause chromosome rearrangements and cancer. The cancer safeguards targeted by EBV extend to the *BRCA* pathway, FA proteins, an SMC5/6 scaffold, JAK-STAT signaling, and the SWI-SNF chromatin remodeling complex. EBV: Epstein-Barr virus; ER: estrogen receptor; FA: Fanconi anemia; MHC: major histocompatibility complex; PDH: pyruvate dehydrogenase; SWI-SNF: switch/sucrose non-fermentable.



The strategy of using bioinformatics to identify markers of "long EBV" may well work for other cancers, multiple sclerosis [103], and other chronic diseases that are currently unexplained. Testing for persistent viral damage in genomes from biopsies is a new method for screening for breast cancer risk. The results may inform further prevention and treatment decisions. Cancer

drug therapy has focused on finding and destroying cancer-driver gene products. The drugs are initially effective, sometimes for long periods, but then stop working. The cycles represented in Figure 10 are an occult, underlying process that can now be evaluated. Cancer treatment generates new clones that do not exist in the original population [104]. The underlying genome damage and EBV scars continually produce new cancer-driver mutations. Some antigens targeted by successful therapy for hematologic malignancies [105], such as DLBCL, may also be effective for breast cancers. The idea that breast cancers and hematologic malignancies can have similar breakpoints and translocation fusions suggests that there may be many more susceptible targets and that there are options to overcome resistance or tolerance [106]. The findings may further stimulate research into other EBV-associated diseases and cancers, leading to better and broader understanding.

Estrogen has been thought to generate the initial chromosome breakpoints leading to translocations in human breast cancer. However, young boys with BL do not produce estrogen from ovaries, yet Figure 3 shows that their malignant B-cells have many breakpoints [68,107] that approximately match breast cancer breakpoints. Normally, aromatase catalyzes the rate-limiting step in estrogen production [108], and aromatase acting on androgens is the primary source of most estrogens in breast tissue [109]. EBV-infected cells lose control of aromatase activity [108]. An EBV-mediated increase in aromatase activity explains why locations of breakpoints (Multimedia Appendix 5) are relatively independent of estrogen receptor status in breast cancer [68] and resemble locations in lymphoid cells (Figures 1-4 and 9). Transcription in response to artificially high estrogen levels created by EBV then induces topoisomerase-mediated DNA breaks. Double-strand break repair genes remove topoisomerase from these complexes, but damage to this process leaves pathological enzyme complexes still bound at a DNA breakpoint [110-112]. As shown in Figure 3, topoisomerase itself may be damaged. In either case, EBV-mediated deregulation of estrogen production, topoisomerase activity, and deaminases then collaborate to cause chromosome breaks and drive breast cancer.

Breast cancer chromosome breakpoints cluster around genes near EBV-binding sites (Figure 4), further suggesting that EBV participates in causing the breaks. The breaks lead to pathogenic chromosome rearrangements because EBV-induced damage forces restoration into error-prone methods by suppressing FA-BRCA pathway intermediates [14,15]. Repairs using the FA-BRCA pathway [113] need chromatin access, which requires the SMC5/6 cohesin complex [114,115]. In one scenario shown in Figure 10, SMC5/6 interacts with a crucial pathway intermediate, the FANCD2-FANCI heterodimer (“D2-I”) [17,116]. EBV variants deplete SMC5/6, preventing FA-BRCA–mediated DNA repairs and leading to chromosomes with too many centromeres. When mitosis pulls apart multicentromere chromosome structures, the forces shatter the chromosome and induce mutation storms [35]. EBV thus threatens a sprawling, interconnected repair system, including the BRCA pathway, FA proteins, an SMC5/6 scaffold, JAK-STAT signaling, and the SWI-SNF chromatin remodeling complex (Figure 10).

Of course, other environmental, genetic, or lifestyle factors also participate in breast cancer development, but EBV infection exacerbates their effects. Genome deficits in EBV-associated cancers and breast cancers interfere with restoring chromosomes from damage due to natural processes and exogenous mutagens. Some of this damage requires repair pathways that are subject to EBV interference.

Evidence underlying the model in Figure 10 has independent support from the literature. For example, viral load is a marker for the extent of cell-free DNA fragmentation [117]. EBV-mediated transformation routinely generates abnormal karyotypes [118]. The binding of EBNA1 sequence variants increases NPC risk and drives EBV lytic gene expression [119,120], which requires EBV-encoded nucleases [121-123]. Other herpesviruses related to EBV share the ability to fragment DNA and subvert DNA repair pathways [124-126]. EBV facilitates its own replication by interfering with signaling pathways that prevent metastasis [29,127-130]. Independent literature supports EBV participation in metastasis and the results shown in Figure 9. NPC has the highest metastatic rate among all head and neck cancers, and the levels of circulating EBV markers are highly predictive [10]. Finding EBV in lymph nodes of patients with NPC or primary cancer at an unknown site helps detect metastasis [131]. NPC patients with ≥500 copies of EBV per mL plasma had significantly higher rates of liver metastasis than patients with lower EBV levels [132]. EBV-infected B-cells and breast cancer cells both have amplified centrosomes (Figure 10), the mitosis-organizing centers that exert structural control over cell division. The EBV protein thymidine kinase takes up residence in the centrosome [133], and another EBV protein, BNRF1, initiates centrosome amplification in infected B-cells [134]. Overduplication of centrosomes confuses chromatid attachments to spindle fibers during mitosis. Chromosomes do not distribute properly into daughter cells, creating mistakes when the genome replicates [134,135]. Neither centrosome amplification nor chromosome fragmentation (chromothripsis) requires large numbers of viral particles or active infection.

Further bioinformatic tests may still add significant additional information. EBV activation brings massive changes to host chromatin methylation and structure [47,51,136]. Breast cancers have hundreds of these changes [34]. Results here further implicate epigenetic effects, so EBV effects on breast cancer epigenetics should be explored in more detail. EBV is implicated in cancers in multiple additional organs, and the methods developed here may help clarify its potential contributions. Predictions based on virus-human interaction structural biology may also be helpful. The ultimate direct test will be whether childhood recipients of an anti-EBV vaccine have reduced breast cancer incidence. If it even approaches the reduction of cervical cancer achieved by the HPV vaccine (up to 94%), a childhood EBV vaccine could effectively prevent many cases of breast cancer.

Limitations

EBV itself creates a limitation because the virus can disappear after causing pathogenic genome damage that allows breast cancer to develop. This transitory virus presence forces the use

of bioinformatics to look for persistent genome damage EBV leaves behind. EBV disappearance questions whether a group of cancers with EBV connections also contains “sporadic” cancers typed as EBV negative. The EBV-negative forms may have merely lost the criteria used to identify EBV infection, but EBV-related genome damage may still remain. Another limitation is that compared to breast cancers, known EBV-linked cancers such as GC, BL, and NPC are less common, so genome sequence data are also less common.

Conclusions

In summary, early childhood immunizations against inactivated EBV or selected EBV gene products may significantly reduce the incidence of breast, ovarian, and other cancers, and potentially unexplained chronic diseases. EBV variants lead to DNA breaks, mitotic abnormalities, and the loss of safeguards that protect against breast cancer and its metastasis. Breast cancer breakpoints cluster around breakpoints in EBV cancers, disrupting genes essential to prevent viral infection and breast cancers. A CRISPR-like region on chromosome 6 sequesters some of the thousands of pieces of EBV sequences in the human genome. The same area of chromosome 6 undergoes variable damage in breast cancer, contributing to the reason not everyone with EBV infection develops cancer. In susceptible people, EBV infection leaves behind pathogenic cancer-associated genome abnormalities (“long EBV”). Clinical implications include improvements in evaluating the chances that cancer will return, increased use of immunotherapy for patients with breast cancer that have active infection, and greater urgency in developing an effective EBV vaccine,.

Data Availability

The primary dataset and calculations that were generated or analyzed during this study are included. Datasets not included are freely available from the original sources or the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Glossary and abbreviations.
[DOCX File, 20 KB - xmed_v6i1e50712_app1.docx]

Multimedia Appendix 2
Most of the calculations used in this work.
[XLSX File, 7644 KB - xmed_v6i1e50712_app2.xlsx]

Multimedia Appendix 3
The exact distances between nasopharyngeal cancer and breast cancer breakpoints on chromosome 1. These breaks gather around a few low valleys that periodically occur across the whole chromosome, but the data points are too numerous to display, making the results difficult to interpret.
[PNG File, 213 KB - xmed_v6i1e50712_app3.png]

Multimedia Appendix 4
Gene functions at breast cancer breakpoints that clustered around breakpoints in EBV-associated cancers (GC, BL, and NPC), and EBNA1-binding sequences reported in the human genome. BL: Burkitt lymphoma; EBNA1: Epstein-Barr virus nuclear antigen 1; EBV: Epstein-Barr virus; NPC: nasopharyngeal cancer.
[DOCX File, 31 KB - xmed_v6i1e50712_app4.docx]

Multimedia Appendix 5
Alternative explanations.
[PNG File, 374 KB - xmed_v6i1e50712_app5.png]

References

1. DeSantis CE, Ma J, Gaudet MM, et al. Breast cancer statistics, 2019. CA Cancer J Clin 2019 Nov;69(6):438-451. [doi: 10.3322/caac.21583] [Medline: 31577379]
2. QuickStats: age-adjusted death rates* for female breast cancer,† by state — National Vital Statistics System, United States, 2019§. MMWR Morb Mortal Wkly Rep 2021 Oct 1;70(39):1391. [doi: 10.15585/mmwr.mm7039a6] [Medline: 34591833]
3. Sarid R, Gao SJ. Viruses and human cancer: from detection to causality. Cancer Lett 2011 Jun 28;305(2):218-227. [doi: 10.1016/j.canlet.2010.09.011] [Medline: 20971551]

4. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer 2002 May;2(5):342-350. [doi: 10.1038/nrc798] [Medline: 12044010]
5. Balfour HHJ, Sifakis F, Sliman JA, Knight JA, Schmeling DO, Thomas W. Age-specific prevalence of Epstein–Barr virus infection among individuals aged 6–19 years in the United States and factors affecting its acquisition. J Infect Dis 2013 Oct 15;208(8):1286-1293. [doi: 10.1093/infdis/jit321] [Medline: 23868878]
6. Huo Q, Zhang N, Yang Q. Epstein-Barr virus infection and sporadic breast cancer risk: a meta-analysis. PLoS One 2012;7(2):e31656. [doi: 10.1371/journal.pone.0031656] [Medline: 22363698]
7. Fina F, Romain S, Ouafik L, et al. Frequency and genome load of Epstein-Barr virus in 509 breast cancers from different geographical areas. Br J Cancer 2001 Mar 23;84(6):783-790. [doi: 10.1054/bjoc.2000.1672] [Medline: 11259092]
8. Hu H, Luo ML, Desmedt C, et al. Epstein-Barr virus infection of mammary epithelial cells promotes malignant transformation. EBioMedicine 2016 Jul;9:148-160. [doi: 10.1016/j.ebiom.2016.05.025] [Medline: 27333046]
9. God JM, Haque A. Burkitt lymphoma: pathogenesis and immune evasion. J Oncol 2010;2010:516047. [doi: 10.1155/2010/516047] [Medline: 20953370]
10. Xu M, Yao Y, Chen H, et al. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. Nat Genet 2019 Jul;51(7):1131-1136. [doi: 10.1038/s41588-019-0436-5] [Medline: 31209392]
11. Bruce JP, To KF, Lui VWY, et al. Whole-genome profiling of nasopharyngeal carcinoma reveals viral-host co-operation in inflammatory NF-κB activation and immune escape. Nat Commun 2021 Jul 7;12(1):4193. [doi: 10.1038/s41467-021-24348-6] [Medline: 34234122]
12. Fountzilas G, Psyrri A, Giannoulatou E, et al. Prevalent somatic BRCA1 mutations shape clinically relevant genomic patterns of nasopharyngeal carcinoma in Southeast Europe. Int J Cancer 2018 Jan 1;142(1):66-80. [doi: 10.1002/ijc.31023] [Medline: 28857155]
13. Devanaboyina M, Kaur J, Whiteley E, et al. NF-κB signaling in tumor pathways focusing on breast and ovarian cancer. Oncol Rev 2022 Oct 3;16:10568. [doi: 10.3389/or.2022.10568] [Medline: 36531159]
14. Lung RWM, Tong JHM, Ip LM, et al. EBV-encoded miRNAs can sensitize nasopharyngeal carcinoma to chemotherapeutic drugs by targeting BRCA1. J Cell Mol Med 2020 Nov;24(22):13523-13535. [doi: 10.1111/jcmm.16007] [Medline: 33074587]
15. Hau PM, Tsao SW. Epstein-Barr virus hijacks DNA damage response transducers to orchestrate its life cycle. Viruses 2017 Nov 16;9(11):341. [doi: 10.3390/v9110341] [Medline: 29144413]
16. Dheekollu J, Wiedmer A, Ayyanathan K, Deakyne JS, Messick TE, Lieberman PM. Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at oriP and viral episome maintenance. Cell 2021 Feb 4;184(3):643-654.e13. [doi: 10.1016/j.cell.2020.12.022] [Medline: 33482082]
17. Rossi F, Helbling-Leclerc A, Kawasumi R, et al. SMC5/6 acts jointly with Fanconi anemia factors to support DNA repair and genome stability. EMBO Rep 2020 Feb 5;21(2):e48222. [doi: 10.15252/embr.201948222] [Medline: 31867888]
18. Yiu SPT, Guo R, Zerbe C, Weekes MP, Gewurz BE. Epstein-Barr virus BNRF1 destabilizes SMC5/6 cohesin complexes to evade its restriction of replication compartments. Cell Rep 2022 Mar 8;38(10):110411. [doi: 10.1016/j.celrep.2022.110411] [Medline: 35263599]
19. Fan Y, Ying H, Wu X, et al. The mutational pattern of homologous recombination (HR)-associated genes and its relevance to the immunotherapeutic response in gastric cancer. Cancer Biol Med 2020 Nov 15;17(4):1002-1013. [doi: 10.20892/j.issn.2095-3941.2020.0089] [Medline: 33299649]
20. Parvin S, Labrada AR, Santiago GE, et al. Novel role of LMO2 in DNA repair control in diffuse large B cell lymphoma. Blood 2016 Dec 2;128(22):776-776. [doi: 10.1182/blood.V128.22.776.776]
21. Busch K, Keller T, Fuchs U, et al. Identification of two distinct MYC breakpoint clusters and their association with various IGH breakpoint regions in the t(8;14) translocations in sporadic Burkitt-lymphoma. Leukemia 2007 Aug;21(8):1739-1751. [doi: 10.1038/sj.leu.2404753] [Medline: 17541401]
22. Chong LC, Ben-Neriah S, Slack GW, et al. High-resolution architecture and partner genes of MYC rearrangements in lymphoma with DLBCL morphology. Blood Adv 2018 Oct 23;2(20):2755-2765. [doi: 10.1182/bloodadvances.2018023572] [Medline: 30348671]
23. Xu J, Chen Y, Olopade OI. MYC and breast cancer. Genes Cancer 2010 Jun;1(6):629-640. [doi: 10.1177/1947601910378691] [Medline: 21779462]
24. Umbreit NT, Zhang CZ, Lynch LD, et al. Mechanisms generating cancer genome complexity from a single cell division error. Science 2020 Apr 17;368(6488):eaba0712. [doi: 10.1126/science.aba0712] [Medline: 32299917]
25. McClintock B. The stability of broken ends of chromosomes in Zea mays. Genetics 1941 Mar;26(2):234-282. [doi: 10.1093/genetics/26.2.234] [Medline: 17247004]
26. Lee JJK, Jung YL, Cheong TC, et al. ERα-associated translocations underlie oncogene amplifications in breast cancer. Nature New Biol 2023 Jun;618(7967):1024-1032. [doi: 10.1038/s41586-023-06057-w] [Medline: 37198482]
27. McAulay KA, Higgins CD, Macsween KF, et al. HLA class I polymorphisms are associated with development of infectious mononucleosis upon primary EBV infection. J Clin Invest 2007 Oct;117(10):3042-3048. [doi: 10.1172/JCI32377] [Medline: 17909631]
28. Aboulaghras S, Khalid A, Makeen HA, et al. Polymorphism of HLA and susceptibility of breast cancer. Front Biosci (Landmark Ed) 2024 Feb 5;29(2):55. [doi: 10.31083/j.fbl2902055] [Medline: 38420797]

29. Yates LR, Knappskog S, Wedge D, et al. Genomic evolution of breast cancer metastasis and relapse. Cancer Cell 2017 Aug 14;32(2):169-184.e7. [doi: 10.1016/j.ccell.2017.07.005] [Medline: 28810143]
30. Friedenson B. Dewey defeats Truman and cancer statistics. J Natl Cancer Inst 2009 Aug 19;101(16):1157. [doi: 10.1093/jnci/djp203] [Medline: 19561316]
31. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature New Biol 2016 Jun 2;534(7605):47-54. [doi: 10.1038/nature17676] [Medline: 27135926]
32. Staaf J, Glodzik D, Bosch A, et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat Med 2019 Oct;25(10):1526-1533. [doi: 10.1038/s41591-019-0582-4] [Medline: 31570822]
33. Nones K, Johnson J, Newell F, et al. Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. Ann Oncol 2019 Jul 1;30(7):1071-1079. [doi: 10.1093/annonc/mdz132] [Medline: 31090900]
34. Batra RN, Lifshitz A, Vidakovic AT, et al. DNA methylation landscapes of 1538 breast cancers reveal a replication-linked clock, epigenomic instability and cis-regulation. Nat Commun 2021 Sep 13;12(1):5406. [doi: 10.1038/s41467-021-25661-w] [Medline: 34518533]
35. Cortés-Ciriano I, Lee JJK, Xi R, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 2020 Mar;52(3):331-341. [doi: 10.1038/s41588-019-0576-7] [Medline: 32025003]
36. Petrucelli N, Daly MB, Pal T. BRCA1- and BRCA2-associated hereditary breast and ovarian cancer. In: Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, editors. GeneReviews: University of Washington, Seattle; 1993. URL: https://www.ncbi.nlm.nih.gov/books/NBK1247/ [accessed 2025-01-09]
37. Lalloo F, Varley J, Moran A, et al. BRCA1, BRCA2 and TP53 mutations in very early-onset breast cancer with associated risks to relatives. Eur J Cancer 2006 May;42(8):1143-1150. [doi: 10.1016/j.ejca.2005.11.032] [Medline: 16644204]
38. Joos S, Falk MH, Lichter P, et al. Variable breakpoints in Burkitt lymphoma cells with chromosomal t(8;14) translocation separate c-myc and the IgH locus up to several hundred kb. Hum Mol Genet 1992 Nov;1(8):625-632. [doi: 10.1093/hmg/1.8.625] [Medline: 1301171]
39. Joos S, Haluska FG, Falk MH, et al. Mapping chromosomal breakpoints of Burkitt's t(8;14) translocations far upstream of c-myc. Cancer Res 1992 Dec 1;52(23):6547-6552. [Medline: 1330296]
40. López C, Kleinheinz K, Aukema SM, et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. Nat Commun 2019 Mar 29;10(1):1459. [doi: 10.1038/s41467-019-08578-3] [Medline: 30926794]
41. Xing R, Zhou Y, Yu J, et al. Whole-genome sequencing reveals novel tandem-duplication hotspots and a prognostic mutational signature in gastric cancer. Nat Commun 2019 May 2;10(1):2037. [doi: 10.1038/s41467-019-09644-6] [Medline: 31048690]
42. Nangalia J, Campbell PJ. Genome sequencing during a patient's journey through cancer. N Engl J Med 2019 Nov 28;381(22):2145-2156. [doi: 10.1056/NEJMra1910138] [Medline: 31774959]
43. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). CSH Protoc 2007 Jul 1;2007(7):pdb.top17. [doi: 10.1101/pdb.top17] [Medline: 21357135]
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990 Oct 5;215(3):403-410. [doi: 10.1016/S0022-2836(05)80360-2] [Medline: 2231712]
45. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol 2000;7(1-2):203-214. [doi: 10.1089/10665270050081478] [Medline: 10890397]
46. Rangwala SH, Kuznetsov A, Ananiev V, et al. Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV). Genome Res 2021 Jan;31(1):159-169. [doi: 10.1101/gr.266932.120] [Medline: 33239395]
47. Kim KD, Tanizawa H, de Leo A, et al. Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. Nat Commun 2020 Feb 13;11(1):877. [doi: 10.1038/s41467-019-14152-8] [Medline: 32054837]
48. Lu F, Wikramasinghe P, Norseen J, et al. Genome-wide analysis of host-chromosome binding sites for Epstein-Barr virus nuclear antigen 1 (EBNA1). Virol J 2010 Oct 7;7:262. [doi: 10.1186/1743-422X-7-262] [Medline: 20929547]
49. Sai Lakshmi S, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. Nucleic Acids Res 2008 Jan;36(Database issue):D173-D177. [doi: 10.1093/nar/gkm696] [Medline: 17881367]
50. Wang J, Zhang P, Lu Y, et al. piRBase: a comprehensive database of piRNA sequences. Nucleic Acids Res 2019 Jan 8;47(D1):D175-D180. [doi: 10.1093/nar/gky1043] [Medline: 30371818]
51. Tang MH, Varadan V, Kamalakaran S, Zhang MQ, Dimitrova N, Hicks J. Major chromosomal breakpoint intervals in breast cancer co-localize with differentially methylated regions. Front Oncol 2012 Dec 27;2:197. [doi: 10.3389/fonc.2012.00197] [Medline: 23293768]
52. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947 Mar;18(1):50-60. [doi: 10.1214/aoms/1177730491]
53. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965 Dec;52(3/4):591-611. [doi: 10.2307/2333709]
54. Kumar R, Nagpal G, Kumar V, Usmani SS, Agrawal P, Raghava GPS. HumCFS: a database of fragile sites in human chromosomes. BMC Genomics 2019 Apr 18;19(Suppl 9):985. [doi: 10.1186/s12864-018-5330-5] [Medline: 30999860]

55. Maccaroni K, Balzano E, Mirimao F, Giunta S, Pelliccia F. Impaired replication timing promotes tissue-specific expression of common fragile sites. Genes (Basel) 2020 Mar 19;11(3):326. [doi: 10.3390/genes11030326] [Medline: 32204553]
56. European Commission: Directorate-General for Research and Innovation. Improving access to and reuse of research results, publications and data for scientific purposes – study to evaluate the effects of the EU copyright framework on research and the effects of potential interventions and to identify and present relevant provisions for research in EU data and digital legislation, with a focus on rights and obligations. : Publications Office of the European Union; 2024 URL: https://data.europa.eu/doi/10.2777/633395 [accessed 2025-01-09]
57. Zheng B, Liu XL, Fan R, et al. The landscape of cell-free HBV integrations and mutations in cirrhosis and hepatocellular carcinoma patients. Clin Cancer Res 2021 Jul 1;27(13):3772-3783. [doi: 10.1158/1078-0432.CCR-21-0002] [Medline: 33947693]
58. Foster KA, Harrington P, Kerr J, et al. Somatic and germline mutations of the BRCA2 gene in sporadic ovarian cancer. Cancer Res 1996 Aug 15;56(16):3622-3625. [Medline: 8705994]
59. Friedenson B. The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. BMC Cancer 2007 Aug 6;7:152. [doi: 10.1186/1471-2407-7-152] [Medline: 17683622]
60. Friedenson B. Comment on 'The incidence of leukaemia in women with BRCA1 and BRCA2 mutations: an International Prospective Cohort Study'. Br J Cancer 2016 Aug 23;115(5):e2. [doi: 10.1038/bjc.2016.192] [Medline: 27459694]
61. Parvin S, Ramirez-Labrada A, Aumann S, et al. LMO2 confers synthetic lethality to PARP inhibition in DLBCL. Cancer Cell 2019 Sep 16;36(3):237-249.e6. [doi: 10.1016/j.ccell.2019.07.007] [Medline: 31447348]
62. Compagno M, Lim WK, Grunn A, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature 2009 Jun 4;459(7247):717-721. [doi: 10.1038/nature07968] [Medline: 19412164]
63. Ceribelli M, Kelly PN, Shaffer AL, et al. Blockade of oncogenic IκB kinase activity in diffuse large B-cell lymphoma by bromodomain and extraterminal domain protein inhibitors. Proc Natl Acad Sci U S A 2014 Aug 5;111(31):11365-11370. [doi: 10.1073/pnas.1411701111] [Medline: 25049379]
64. Montes-Moreno S, Odqvist L, Diaz-Perez JA, et al. EBV-positive diffuse large B-cell lymphoma of the elderly is an aggressive post-germinal center B-cell neoplasm characterized by prominent nuclear factor-kB activation. Mod Pathol 2012 Jul;25(7):968-982. [doi: 10.1038/modpathol.2012.52] [Medline: 22538516]
65. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. Nature 2018 Mar 15;555(7696):321-327. [doi: 10.1038/nature25480] [Medline: 29489754]
66. Li JSZ, Abbasi A, Kim DH, Lippman SM, Alexandrov LB, Cleveland DW. Chromosomal fragile site breakage by EBV-encoded EBNA1 at clustered repeats. Nature 2023 Apr;616(7957):504-509. [doi: 10.1038/s41586-023-05923-x] [Medline: 37046091]
67. Hui KF, Chan TF, Yang W, et al. High risk Epstein‐Barr virus variants characterized by distinct polymorphisms in the EBER locus are strongly associated with nasopharyngeal carcinoma. Int J Cancer 2019 Jun 15;144(12):3031-3042. [doi: 10.1002/ijc.32049] [Medline: 30536939]
68. Friedenson B. Evidence of lesions from Epstein-Barr virus infection in human breast cancer genomes. medRxiv. Preprint posted online on Jun 26, 2024. [doi: 10.1101/2024.06.24.24309410]
69. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. Science 2022 Apr;376(6588):44-53. [doi: 10.1126/science.abj6987] [Medline: 35357919]
70. Luo WJ, Takakuwa T, Ham MF, et al. Epstein-Barr virus is integrated between REL and BCL-11A in American Burkitt lymphoma cell line (NAB-2). Lab Invest 2004 Sep;84(9):1193-1199. [doi: 10.1038/labinvest.3700152] [Medline: 15241441]
71. Tsao SW, Tsang CM, Lo KW. Epstein-Barr virus infection and nasopharyngeal carcinoma. Philos Trans R Soc Lond B Biol Sci 2017 Oct 19;372(1732):20160270. [doi: 10.1098/rstb.2016.0270] [Medline: 28893937]
72. Scott RS. Epstein-Barr virus: a master epigenetic manipulator. Curr Opin Virol 2017 Oct;26:74-80. [doi: 10.1016/j.coviro.2017.07.017] [Medline: 28780440]
73. Schott G, Garcia-Blanco MA. MHC class III RNA binding proteins and immunity. RNA Biol 2021 May;18(5):640-646. [doi: 10.1080/15476286.2020.1860388] [Medline: 33280511]
74. Zhou D, Lai M, Luo A, Yu CY. An RNA metabolism and surveillance quartet in the major histocompatibility complex. Cells 2019 Aug 30;8(9):1008. [doi: 10.3390/cells8091008] [Medline: 31480283]
75. Fernandes MT, Dejardin E, dos Santos NR. Context-dependent roles for lymphotoxin-β receptor signaling in cancer development. Biochim Biophys Acta 2016 Apr;1865(2):204-219. [doi: 10.1016/j.bbcan.2016.02.005] [Medline: 26923876]
76. Shiina T, Blancher A, Inoko H, Kulski JK. Comparative genomics of the human, macaque and mouse major histocompatibility complex. Immunology 2017 Feb;150(2):127-138. [doi: 10.1111/imm.12624] [Medline: 27395034]
77. Aissani B, Boehme AK, Wiener HW, Shrestha S, Jacobson LP, Kaslow RA. SNP screening of central MHC-identified HLA-DMB as a candidate susceptibility gene for HIV-related Kaposi's sarcoma. Genes Immun 2014 Sep;15(6):424-429. [doi: 10.1038/gene.2014.42] [Medline: 25008864]
78. Zhuang J, Huo Q, Yang F, Xie N. Perspectives on the role of histone modification in breast cancer progression and the advanced technological tools to study epigenetic determinants of metastasis. Front Genet 2020 Oct 29;11:603552. [doi: 10.3389/fgene.2020.603552] [Medline: 33193750]

79. Bayona-Feliu A, Barroso S, Muñoz S, Aguilera A. The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription–replication conflicts. Nat Genet 2021 Jul;53(7):1050-1063. [doi: 10.1038/s41588-021-00867-2] [Medline: 33986538]
80. Harrod A, Lane KA, Downs JA. The role of the SWI/SNF chromatin remodelling complex in the response to DNA double strand breaks. DNA Repair (Amst) 2020 Sep;93:102919. [doi: 10.1016/j.dnarep.2020.102919] [Medline: 33087260]
81. Andrades A, Peinado P, Alvarez-Perez JC, et al. SWI/SNF complexes in hematological malignancies: biological implications and therapeutic opportunities. Mol Cancer 2023 Feb 21;22(1):39. [doi: 10.1186/s12943-023-01736-8] [Medline: 36810086]
82. Lu J. The Warburg metabolism fuels tumor metastasis. Cancer Metastasis Rev 2019 Jun;38(1-2):157-164. [doi: 10.1007/s10555-019-09794-5] [Medline: 30997670]
83. Chen H, Wu J, Zhang Z, et al. Association between BRCA status and triple-negative breast cancer: a meta-analysis. Front Pharmacol 2018 Aug 21;9:909. [doi: 10.3389/fphar.2018.00909] [Medline: 30186165]
84. Evans DG, Lalloo F, Howell S, Verhoef S, Woodward ER, Howell A. Low prevalence of HER2 positivity amongst BRCA1 and BRCA2 mutation carriers and in primary BRCA screens. Breast Cancer Res Treat 2016 Feb;155(3):597-601. [doi: 10.1007/s10549-016-3697-z] [Medline: 26888723]
85. Locy H, Verhulst S, Cools W, et al. Assessing tumor-infiltrating lymphocytes in breast cancer: a proposal for combining immunohistochemistry and gene expression analysis to refine scoring. Front Immunol 2022 Feb 11;13:794175. [doi: 10.3389/fimmu.2022.794175] [Medline: 35222378]
86. Takada K, Kashiwagi S, Asano Y, et al. Prediction of distant metastatic recurrence by tumor-infiltrating lymphocytes in hormone receptor-positive breast cancer. BMC Womens Health 2021 May 29;21(1):225. [doi: 10.1186/s12905-021-01373-7] [Medline: 34051785]
87. Meier UC, Cipian RC, Karimi A, Ramasamy R, Middeldorp JM. Cumulative roles for Epstein-Barr virus, human endogenous retroviruses, and human herpes virus-6 in driving an inflammatory cascade underlying MS pathogenesis. Front Immunol 2021 Nov 1;12:757302. [doi: 10.3389/fimmu.2021.757302] [Medline: 34790199]
88. Mameli G, Poddighe L, Mei A, et al. Expression and activation by Epstein Barr virus of human endogenous retroviruses-W in blood cells and astrocytes: inference for multiple sclerosis. PLoS One 2012;7(9):e44991. [doi: 10.1371/journal.pone.0044991] [Medline: 23028727]
89. Lawson JS, Glenn WK. Evidence for a causal role by mouse mammary tumour-like virus in human breast cancer. NPJ Breast Cancer 2019 Nov 7;5:40. [doi: 10.1038/s41523-019-0136-4] [Medline: 31728407]
90. Dudley JP, Golovkina TV, Ross SR. Lessons learned from mouse mammary tumor virus in animal models. ILAR J 2016;57(1):12-23. [doi: 10.1093/ilar/ilv044] [Medline: 27034391]
91. Li Y, Roberts ND, Wala JA, et al. Patterns of somatic structural variation in human cancer genomes. Nature 2020 Feb;578(7793):112-121. [doi: 10.1038/s41586-019-1913-9] [Medline: 32025012]
92. Ganapathiraju MK, Subramanian S, Chaparala S, Karunakaran KB. A reference catalog of DNA palindromes in the human genome and their variations in 1000 Genomes. Hum Genome Var 2020 Nov 20;7(1):40. [doi: 10.1038/s41439-020-00127-5] [Medline: 33298903]
93. Rawlins DR, Milman G, Hayward SD, Hayward GS. Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. Cell 1985 Oct;42(3):859-868. [doi: 10.1016/0092-8674(85)90282-x] [Medline: 2996781]
94. Bochkarev A, Barwell JA, Pfuetzner RA, Bochkareva E, Frappier L, Edwards AM. Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA. Cell 1996 Mar 8;84(5):791-800. [doi: 10.1016/s0092-8674(00)81056-9] [Medline: 8625416]
95. Altemose N, Logsdon GA, Bzikadze AV, et al. Complete genomic and epigenetic maps of human centromeres. Science 2022 Apr;376(6588):eabl4178. [doi: 10.1126/science.abl4178] [Medline: 35357911]
96. Grande BM, Gerhard DS, Jiang A, et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. Blood 2019 Mar 21;133(12):1313-1324. [doi: 10.1182/blood-2018-09-871418] [Medline: 30617194]
97. Kim TM, Xi R, Luquette LJ, Park RW, Johnson MD, Park PJ. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. Genome Res 2013 Feb;23(2):217-227. [doi: 10.1101/gr.140301.112] [Medline: 23132910]
98. Dittmer DP, Hilscher CJ, Gulley ML, Yang EV, Chen M, Glaser R. Multiple pathways for Epstein-Barr virus episome loss from nasopharyngeal carcinoma. Int J Cancer 2008 Nov 1;123(9):2105-2112. [doi: 10.1002/ijc.23685] [Medline: 18688856]
99. Li Q, Cohen JI. Epstein-Barr virus and the human leukocyte antigen complex. Curr Clin Microbiol Rep 2019 Sep;6(3):175-181. [doi: 10.1007/s40588-019-00120-9] [Medline: 33094090]
100. Lei PJ, Pereira ER, Andersson P, et al. Cancer cell plasticity and MHC-II–mediated immune tolerance promote breast cancer metastasis to lymph nodes. J Exp Med 2023 Sep 4;220(9):e20221847. [doi: 10.1084/jem.20221847] [Medline: 37341991]
101. Park IA, Hwang SH, Song IH, et al. Expression of the MHC class II in triple-negative breast cancer is associated with tumor-infiltrating lymphocytes and interferon signaling. PLoS One 2017 Aug 17;12(8):e0182786. [doi: 10.1371/journal.pone.0182786] [Medline: 28817603]

102. Song IH, Kim YA, Heo SH, et al. The association of estrogen receptor activity, interferon signaling, and MHC class I expression in breast cancer. Cancer Res Treat 2022 Oct;54(4):1111-1120. [doi: 10.4143/crt.2021.1017] [Medline: 34942685]
103. Bjornevik K, Cortese M, Healy BC, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science 2022 Jan 21;375(6578):296-301. [doi: 10.1126/science.abj8222] [Medline: 35025605]
104. Aissa AF, Islam ABMMK, Ariss MM, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun 2021 Mar 12;12(1):1628. [doi: 10.1038/s41467-021-21884-z] [Medline: 33712615]
105. Cao Y, Efetov SK, He M, et al. Updated clinical perspectives and challenges of chimeric antigen receptor-T cell therapy in colorectal cancer and invasive breast cancer. Arch Immunol Ther Exp (Warsz) 2023 Aug 11;71(1):19. [doi: 10.1007/s00005-023-00684-x] [Medline: 37566162]
106. Upton R, Banuelos A, Feng D, et al. Combining CD47 blockade with trastuzumab eliminates HER2-positive breast cancer cells and overcomes trastuzumab tolerance. Proc Natl Acad Sci U S A 2021 Jul 20;118(29):e2026849118. [doi: 10.1073/pnas.2026849118] [Medline: 34257155]
107. Burkhardt B, Michgehl U, Rohde J, et al. Clinical relevance of molecular characteristics in Burkitt lymphoma differs according to age. Nat Commun 2022 Jul 6;13(1):3881. [doi: 10.1038/s41467-022-31355-8] [Medline: 35794096]
108. Dochi H, Kondo S, Murata T, et al. Estrogen induces the expression of EBV lytic protein ZEBRA, a marker of poor prognosis in nasopharyngeal carcinoma. Cancer Sci 2022 Aug;113(8):2862-2877. [doi: 10.1111/cas.15440] [Medline: 35633182]
109. Zhu BT, Conney AH. Functional role of estrogen metabolism in target cells: review and perspectives. Carcinogenesis 1998 Jan 1;19(1):1-27. [doi: 10.1093/carcin/19.1.1] [Medline: 9472688]
110. Pommier Y, Nussenzweig A, Takeda S, Austin C. Human topoisomerases and their roles in genome stability and organization. Nat Rev Mol Cell Biol 2022 Jun;23(6):407-427. [doi: 10.1038/s41580-022-00452-3] [Medline: 35228717]
111. Sasanuma H, Tsuda M, Morimoto S, et al. BRCA1 ensures genome integrity by eliminating estrogen-induced pathological topoisomerase II-DNA complexes. Proc Natl Acad Sci U S A 2018 Nov 6;115(45):E10642-E10651. [doi: 10.1073/pnas.1803177115] [Medline: 30352856]
112. Manguso N, Kim M, Joshi N, et al. TDP2 is a regulator of estrogen-responsive oncogene expression. NAR Cancer 2024 Apr 8;6(2):zcae016. [doi: 10.1093/narcan/zcae016] [Medline: 38596431]
113. de Piccoli G, Cortes-Ledesma F, Ira G, et al. Smc5-Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. Nat Cell Biol 2006 Sep;8(9):1032-1034. [doi: 10.1038/ncb1466] [Medline: 16892052]
114. Tapia-Alveal C, Lin SJ, O'Connell MJ. Functional interplay between cohesin and Smc5/6 complexes. Chromosoma 2014 Oct;123(5):437-445. [doi: 10.1007/s00412-014-0474-9] [Medline: 24981336]
115. Irwan ID, Cullen BR. The SMC5/6 complex: an emerging antiviral restriction factor that can silence episomal DNA. PLoS Pathog 2023 Mar 2;19(3):e1011180. [doi: 10.1371/journal.ppat.1011180] [Medline: 36862666]
116. Agashe S, Joseph CR, Reyes TAC, et al. Smc5/6 functions with Sgs1-Top3-Rmi1 to complete chromosome replication at natural pause sites. Nat Commun 2021 Apr 8;12(1):2111. [doi: 10.1038/s41467-021-22217-w] [Medline: 33833229]
117. Truszewska A, Wirkowska A, Gala K, et al. EBV load is associated with cfDNA fragmentation and renal damage in SLE patients. Lupus 2021 Jul;30(8):1214-1225. [doi: 10.1177/09612033211010339] [Medline: 33866897]
118. Volleth M, Zenker M, Joksic I, Liehr T. Long-term culture of EBV-induced human lymphoblastoid cell lines reveals chromosomal instability. J Histochem Cytochem 2020 Apr;68(4):239-251. [doi: 10.1369/0022155420910113] [Medline: 32108534]
119. Dheekollu J, Malecka K, Wiedmer A, et al. Carcinoma-risk variant of EBNA1 deregulates Epstein-Barr virus episomal latency. Oncotarget 2017 Jan 31;8(5):7248-7264. [doi: 10.18632/oncotarget.14540] [Medline: 28077791]
120. Buisson M, Géoui T, Flot D, et al. A bridge crosses the active-site canyon of the Epstein-Barr virus nuclease with DNase and RNase activities. J Mol Biol 2009 Aug 28;391(4):717-728. [doi: 10.1016/j.jmb.2009.06.034] [Medline: 19538972]
121. Wu CC, Liu MT, Chang YT, et al. Epstein-Barr virus DNase (BGLF5) induces genomic instability in human epithelial cells. Nucleic Acids Res 2010 Apr;38(6):1932-1949. [doi: 10.1093/nar/gkp1169] [Medline: 20034954]
122. Chiu SH, Wu MC, Wu CC, et al. Epstein-Barr virus BALF3 has nuclease activity and mediates mature virion production during the lytic cycle. J Virol 2014 May;88(9):4962-4975. [doi: 10.1128/JVI.00063-14] [Medline: 24554665]
123. Chiu SH, Wu CC, Fang CY, et al. Epstein-Barr virus BALF3 mediates genomic instability and progressive malignancy in nasopharyngeal carcinoma. Oncotarget 2014 Sep 30;5(18):8583-8601. [doi: 10.18632/oncotarget.2323] [Medline: 25261366]
124. O'Neill FJ, Miles CP. Chromosome changes in human cells induced by herpes simplex, types 1 and 2. Nature 1969 Aug 23;223(5208):851-852. [doi: 10.1038/223851a0] [Medline: 4307971]
125. Mertens ME, Knipe DM. Herpes simplex virus 1 manipulates host cell antiviral and proviral DNA damage responses. mBio 2021 Feb 9;12(1):e03552-20. [doi: 10.1128/mBio.03552-20] [Medline: 33563816]
126. Schumacher AJ, Mohni KN, Kan Y, Hendrickson EA, Stark JM, Weller SK. The HSV-1 exonuclease, UL12, stimulates recombination by a single strand annealing mechanism. PLoS Pathog 2012;8(8):e1002862. [doi: 10.1371/journal.ppat.1002862] [Medline: 22912580]
127. Ezeonwumelu IJ, Garcia-Vidal E, Ballana E. JAK-STAT pathway: a novel target to tackle viral infections. Viruses 2021 Nov 27;13(12):2379. [doi: 10.3390/v13122379] [Medline: 34960648]

128. Jangra S, Bharti A, Lui WY, et al. Suppression of JAK-STAT signaling by Epstein-Barr virus tegument protein BGLF2 through recruitment of SHP1 phosphatase and promotion of STAT2 degradation. J Virol 2021 Sep 27;95(20):e0102721. [doi: 10.1128/JVI.01027-21] [Medline: 34319780]
129. Wu DY, Krumm A, Schubach WH. Promoter-specific targeting of human SWI-SNF complex by Epstein-Barr virus nuclear protein 2. J Virol 2000 Oct;74(19):8893-8903. [doi: 10.1128/jvi.74.19.8893-8903.2000] [Medline: 10982332]
130. Su MT, Wang YT, Chen YJ, Lin SF, Tsai CH, Chen MR. The SWI/SNF chromatin regulator BRG1 modulates the transcriptional regulatory activity of the Epstein-Barr virus DNA polymerase processivity factor BMRF1. J Virol 2017 Apr 13;91(9):e02114-16. [doi: 10.1128/JVI.02114-16] [Medline: 28228591]
131. Nakao K, Yuge T, Mochiki M, Nibu KI, Sugasawa M. Detection of Epstein-Barr virus in metastatic lymph nodes of patients with nasopharyngeal carcinoma and a primary unknown carcinoma. Arch Otolaryngol Head Neck Surg 2003 Mar;129(3):338-340. [doi: 10.1001/archotol.129.3.338] [Medline: 12622545]
132. Tao D, Zhang N, Huang Q, et al. Association of Epstein-Barr virus infection with peripheral immune parameters and clinical outcome in advanced nasopharyngeal carcinoma. Sci Rep 2020 Dec 15;10(1):21976. [doi: 10.1038/s41598-020-78892-0] [Medline: 33319825]
133. Gill MB, Kutok JL, Fingeroth JD. Epstein-Barr virus thymidine kinase is a centrosomal resident precisely localized to the periphery of centrioles. J Virol 2007 Jun 15;81(12):6523-6535. [doi: 10.1128/JVI.00147-07] [Medline: 17428875]
134. Shumilov A, Tsai MH, Schlosser YT, et al. Epstein-Barr virus particles induce centrosome amplification and chromosomal instability. Nat Commun 2017 Feb 10;8:14257. [doi: 10.1038/ncomms14257] [Medline: 28186092]
135. Lingle WL, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL. Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. Proc Natl Acad Sci U S A 1998 Mar 17;95(6):2950-2955. [doi: 10.1073/pnas.95.6.2950] [Medline: 9501196]
136. Buschle A, Mrozek-Gorska P, Cernilogar FM, et al. Epstein-Barr virus inactivates the transcriptome and disrupts the chromatin architecture of its host cell in the first phase of lytic reactivation. Nucleic Acids Res 2021 Apr 6;49(6):3217-3241. [doi: 10.1093/nar/gkab099] [Medline: 33675667]

Abbreviations

BL: Burkitt lymphoma
BLAST: Basic Local Alignment Search Tool
bp: base pairs
COSMIC: Catalog of Somatic Mutations in Cancer
CRISPR: clustered regularly interspaced short palindromic repeats
DLBCL: diffuse large B-cell lymphoma
EBNA1: Epstein-Barr virus nuclear antigen 1
EBV: Epstein-Barr virus
FA: Fanconi anemia
GC: gastric cancer
HER2: human epidermal growth factor receptor 2
HERV: human endogenous retrovirus
HIV1: human immunodeficiency virus type 1
HLA: human leukocyte antigen
HPV: human papillomavirus
JAK: Janus kinase
MHC: major histocompatibility complex
NF-κB: nuclear factor–κB
NPC: nasopharyngeal cancer
piRNA: Piwi-interacting RNA
SWI-SNF : switch/sucrose non-fermentable
TIL: tumor-infiltrating lymphocyte

Edited by A Schwartz; submitted 10.07.23; peer-reviewed by Anonymous, Anonymous; revised version received 19.11.24; accepted 20.11.24; published 29.01.25.

Please cite as:
Friedenson B
Identifying Safeguards Disabled by Epstein-Barr Virus Infections in Genomes From Patients With Breast Cancer: Chromosomal Bioinformatics Analysis
JMIRx Med 2025;6:e50712
URL: https://xmed.jmir.org/2025/1/e50712
doi:10.2196/50712

Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection

Mahesh Vaijainthymala Krishnamoorthy, BE

Stelmith, LLC, 2333 Aberdeen Pl, Carollton, TX, United States

Corresponding Author:
Mahesh Vaijainthymala Krishnamoorthy, BE
Stelmith, LLC, 2333 Aberdeen Pl, Carollton, TX, United States

Related Articles:

Companion article: https://arxiv.org/abs/2410.17459v1

Companion article: https://med.jmirx.org/2025/1/e72523

Companion article: https://med.jmirx.org/2025/1/e72525

Companion article: https://med.jmirx.org/2025/1/e72527

Abstract

Background: The increasing integration of artificial intelligence (AI) systems into critical societal sectors has created an urgent demand for robust privacy-preserving methods. Traditional approaches such as differential privacy and homomorphic encryption often struggle to maintain an effective balance between protecting sensitive information and preserving data utility for AI applications. This challenge has become particularly acute as organizations must comply with evolving AI governance frameworks while maintaining the effectiveness of their AI systems.

Objective: This paper aims to introduce and validate data obfuscation through latent space projection (LSP), a novel privacy-preserving technique designed to enhance AI governance and ensure responsible AI compliance. The primary goal is to develop a method that can effectively protect sensitive data while maintaining essential features necessary for AI model training and inference, thereby addressing the limitations of existing privacy-preserving approaches.

Methods: We developed LSP using a combination of advanced machine learning techniques, specifically leveraging autoencoder architectures and adversarial training. The method projects sensitive data into a lower-dimensional latent space, where it separates sensitive from nonsensitive information. This separation enables precise control over privacy-utility trade-offs. We validated LSP through comprehensive experiments on benchmark datasets and implemented 2 real-world case studies: a health care application focusing on cancer diagnosis and a financial services application analyzing fraud detection.

Results: LSP demonstrated superior performance across multiple evaluation metrics. In image classification tasks, the method achieved 98.7% accuracy while maintaining strong privacy protection, providing 97.3% effectiveness against sensitive attribute inference attacks. This performance significantly exceeded that of traditional anonymization and privacy-preserving methods. The real-world case studies further validated LSP's effectiveness, showing robust performance in both health care and financial applications. Additionally, LSP demonstrated strong alignment with global AI governance frameworks, including the General Data Protection Regulation, the California Consumer Privacy Act, and the Health Insurance Portability and Accountability Act.

Conclusions: LSP represents a significant advancement in privacy-preserving AI, offering a promising approach to developing AI systems that respect individual privacy while delivering valuable insights. By embedding privacy protection directly within the machine learning pipeline, LSP contributes to key principles of fairness, transparency, and accountability. Future research directions include developing theoretical privacy guarantees, exploring integration with federated learning systems, and enhancing latent space interpretability. These developments position LSP as a crucial tool for advancing ethical AI practices and ensuring responsible technology deployment in privacy-sensitive domains.

(JMIRx Med 2025;6:e70100) doi:10.2196/70100

KEYWORDS

privacy-preserving AI; latent space projection; data obfuscation; AI governance; machine learning privacy; differential privacy; k-anonymity; HIPAA; GDPR; compliance; data utility; privacy-utility trade-off; responsible AI; medical imaging privacy; secure data sharing; artificial intelligence; General Data Protection Regulation; Health Insurance Portability and Accountability Act

Introduction

Background

The rapid advancement and widespread adoption of artificial intelligence (AI) across critical sectors of society have ushered in an era of unprecedented data analysis and decision-making capabilities. From health care diagnostics to financial fraud detection, AI systems are processing increasingly large volumes of sensitive personal data. However, this progress has been accompanied by growing concerns about privacy, data protection, and the potential misuse of personal information.

The tension between leveraging data for AI advancements and protecting individual privacy has become a central challenge in the field of AI governance. Traditional approaches to data privacy, such as anonymization and differential privacy, often struggle to balance the trade-off between privacy protection and data utility. As AI systems become more sophisticated, there is an urgent need for novel privacy-preserving techniques that can protect sensitive information without significantly compromising the performance of AI models.

In this research, we introduce data obfuscation through latent space projection (LSP), a novel privacy-preserving technique designed to address these challenges. LSP leverages recent advancements in representation learning and adversarial training to create a privacy-preserving data transformation pipeline. By projecting raw data into a latent space and then reconstructing it with carefully controlled information loss, we aim to obfuscate sensitive attributes while preserving the overall structure and relationships within the data that are crucial for AI model performance.

This research makes several significant contributions to the field of privacy-preserving machine learning. At the core of this work, we develop and present a comprehensive latent space projection framework, providing detailed insights into its theoretical underpinnings, architectural design, and practical implementation considerations. We advance the field's measurement capabilities by introducing innovative metrics specifically designed to evaluate the critical balance between privacy protection and data utility in latent space representations. Through rigorous experimentation on established benchmark datasets, we demonstrate that LSP consistently outperforms traditional privacy-preserving approaches across multiple performance dimensions.

To bridge the gap between theory and practice, we showcase LSP's real-world effectiveness through 2 critical case studies in highly sensitive domains: cancer diagnosis and financial fraud detection. Understanding the practical constraints of deployment, we conduct thorough analyses of LSP's operational characteristics, including latency and computational resource requirements. Finally, we explore the broader implications of our work, examining how LSP contributes to the responsible development of AI systems and aligns with emerging global AI governance frameworks, providing a foundation for future privacy-preserving AI applications.

The Privacy Challenge in AI

The exponential growth of data and the increasing sophistication of AI models have led to significant advancements in various fields. However, this progress has also raised critical privacy concerns [1]. AI models, particularly deep learning architectures, often require vast amounts of data to achieve high performance. This data frequently contains sensitive personal information, ranging from medical records to financial transactions.

The potential for privacy breaches in AI systems is multifaceted and detailed in the following sections.

Data Breaches

Large datasets used for AI training are attractive targets for cyberattacks, potentially exposing the sensitive information of millions of individuals[2,3].

Model Inversion Attacks

Sophisticated attacks can potentially reconstruct training data from model parameters, compromising the privacy of individuals in the training set [4].

Membership Inference

These attacks aim to determine whether a particular data point was used in training a model, which can reveal sensitive information about individuals [5].

Attribute Inference

Even when direct identifiers are removed, AI models may inadvertently learn and expose sensitive attributes of individuals in their training data [6].

Unintended Memorization

Neural networks have been shown to sometimes memorize specific data points from their training set, potentially exposing sensitive information during inference [7].

These privacy risks are not merely theoretical. High-profile incidents of privacy breaches and misuse of personal data have eroded public trust in AI systems and raised regulatory scrutiny. Consequently, there is an urgent need for robust privacy-preserving techniques that can mitigate these risks while allowing AI to deliver its potential benefits to society.

Existing Privacy-Preserving Techniques

Several approaches have been developed to address privacy concerns in AI.

K-Anonymity

Introduced by Sweeney [8], k-anonymity ensures that each record in a dataset is indistinguishable from at least k-1 other records with respect to certain identifying attributes. Although

effective for simple datasets, k-anonymity struggles with high-dimensional data common in modern AI applications.

Differential Privacy

Developed by Dwork et al [9], differential privacy provides a formal framework for quantifying and limiting the privacy risk of statistical queries on datasets. It has been successfully applied to various machine learning algorithms [10,11] but often introduces a significant trade-off between privacy and model utility.

Homomorphic Encryption

This technique allows computations to be performed on encrypted data without decryption [12]. Although providing strong privacy guarantees, homomorphic encryption incurs substantial computational overhead, making it impractical for many real-time AI applications.

Federated Learning

Proposed by McMahan et al [13], federated learning allows models to be trained on decentralized data without directly sharing raw information. However, it can still be vulnerable to certain types of privacy attacks and faces challenges in scenarios requiring centralized data analysis.

Synthetic Data Generation

Techniques like differentially private generative adversarial networks (GANs) [14] aim to generate synthetic datasets that preserve statistical properties of the original data while providing privacy guarantees. However, these methods often struggle to capture complex relationships present in real-world data.

Although each of these approaches has its merits, they all face limitations when applied to the complex, high-dimensional datasets typical in modern AI applications. Many struggle to provide strong privacy guarantees without significantly degrading model performance or incurring prohibitive computational costs.

The Promise of Latent Space Approaches

Recent advancements in representation learning, particularly in the field of deep learning, have opened new avenues for privacy-preserving data analysis [15]. Latent space models, such as autoencoders and variational autoencoders [16], have demonstrated a remarkable ability to learn compact, abstract representations of complex data.

Latency Characteristics

LSP's latency profile can be broken down into three main components: (1) encoding latency (the time taken to project input data into the latent space), (2) processing latency (the time required to perform operations, eg, machine learning tasks, in the latent space), and (3) decoding latency (the time needed to reconstruct data from the latent space, if required).

Performance Optimization Characteristics

These latent representations offer several potential advantages for privacy-preserving AI. Several optimizations contribute to LSP's improved latency and overall performance:

1. Dimensionality reduction: By projecting data into a lower-dimensional latent space, LSP reduces the computational complexity of subsequent operations, so irrelevant or sensitive features can be naturally obscured. This is particularly beneficial for high-dimensional data like images or complex time series.
2. Parallel processing: The encoder and decoder networks in LSP can leverage the parallel processing capabilities of modern GPUs, significantly speeding up the projection and reconstruction processes.
3. Caching mechanisms: For scenarios where the same data are processed multiple times, LSP implementations can cache latent representations, eliminating the need for repeated encoding.
4. Model compression: Techniques such as pruning and quantization can be applied to the LSP networks, reducing their size, and improving inference speed without significantly impacting privacy or utility.
5. Adaptive computation: LSP can be implemented with adaptive computation techniques, where the depth or width of the network is dynamically adjusted based on the complexity of the input, further optimizing performance.
6. Disentanglement: Advanced techniques in representation learning aim to disentangle different factors of variation in the data, potentially allowing for selective obfuscation of sensitive attributes.
7. Nonlinear transformations: The complex, nonlinear mappings learned by deep neural networks can potentially create representations that are difficult to invert without knowledge of the encoding process.
8. Compatibility with deep learning: Latent space approaches integrate naturally with deep learning architectures, allowing for end-to-end privacy-preserving AI pipelines.

Building on these insights, our proposed LSP technique aims to leverage the power of latent space representations to create a robust, flexible framework for privacy-preserving AI. By combining ideas from representation learning, adversarial training, and information theory, LSP seeks to overcome the limitations of existing approaches and provide a more effective solution to the privacy challenges in modern AI systems.

Related Work

Privacy-preserving techniques in AI have garnered significant attention, particularly as regulations such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) come into force. Existing methods provide foundational solutions but have limitations when applied to large-scale data systems.

Differential Privacy

Differential privacy, introduced by Dwork et al [17], is a method that adds calibrated noise to datasets or model outputs to obscure individual data points while preserving the overall distribution. Despite its utility, differential privacy often introduces trade-offs between privacy and model accuracy, particularly when applied to complex, high-dimensional data [18].

Homomorphic Encryption

Homomorphic encryption allows computations to be performed on encrypted data without decrypting it [12]. Although this approach is highly secure, its computational overhead makes it impractical for large-scale machine learning models that require real-time processing or high-volume datasets [19].

Federated Learning

Federated learning, proposed by McMahan et al [13], ensures that raw data remains decentralized, with models trained on local devices instead of centralized servers. However, this technique is not immune to privacy risks, as model gradients or weights exchanged between devices can still leak sensitive information [20,21].

Generative Models for Privacy

Recent work has explored the use of generative models, such as GANs, for creating synthetic data that preserves privacy [22]. Although promising, these approaches often struggle with mode collapse and may not fully capture the complexity of real-world data distributions.

LSP builds upon these existing approaches while addressing their limitations. By learning privacy-preserving latent representations, LSP aims to provide a more flexible and efficient solution for data obfuscation that can be applied across various domains and AI tasks.

Methods

Data Obfuscation Through LSP

In this section, we present the details of our LSP framework for privacy-preserving data obfuscation. We begin by outlining the key principles behind LSP, then describe the network architecture and training procedure.

Principles of LSP

The core idea behind LSP is to transform raw data into a latent space where sensitive information is obscured, yet essential features for downstream AI tasks are retained. This is achieved through the following key principles.

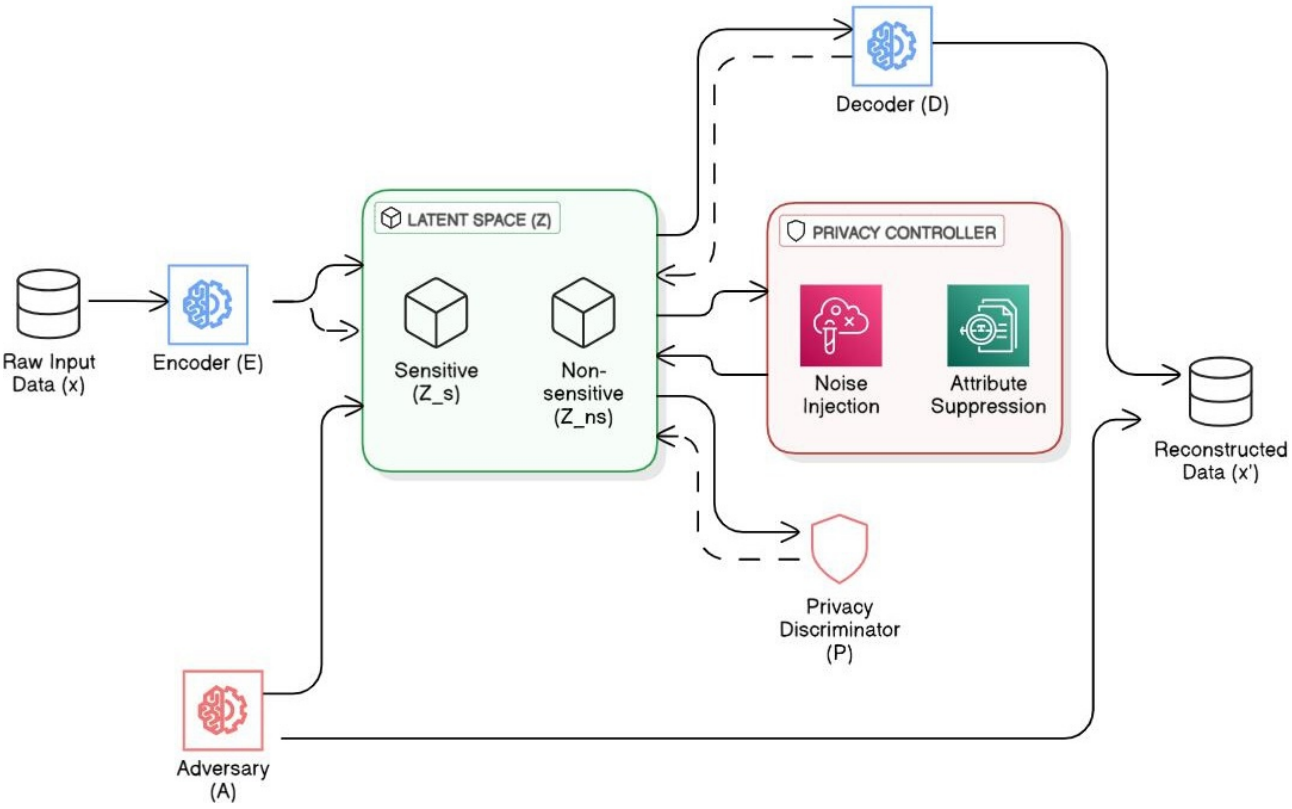
- Feature preservation: The latent representation should maintain sufficient information for relevant AI tasks, ensuring high utility of the obfuscated data.
- Adversarial privacy: We employ adversarial training to make it difficult for an attacker to recover sensitive information from the latent representation.
- Task-agnostic design: The LSP framework is designed to be adaptable to various data types and downstream tasks without requiring significant modifications.

Network Architecture

Figure 1 depicts the flow of data through the LSP framework. The input data x is first passed through the encoder network E, which projects it into a latent space representation z. This latent representation is then processed by the decoder network D to reconstruct the input, producing x'. Simultaneously, the privacy discriminator P attempts to extract sensitive information s from the latent representation z. The framework is trained adversarial to optimize the trade-off between reconstruction accuracy and privacy protection.

The LSP framework consists of three main components: an encoder network, a decoder network, and a privacy discriminator. These components work together to create privacy-preserving latent representations of the input data. Figure 1 illustrates the overall architecture of the LSP framework.

Figure 1. Latent space projection system architecture (network diagram).



Encoder Network

The encoder network E ($X \rightarrow Z$) maps the input data $x \in X$ to a latent representation $z \in Z$. We implement E as a deep neural network with an architecture tailored to the specific data type.

For image data, the encoder architecture uses a progressive series of convolutional layers with expanding filter sizes, beginning at 32 and scaling up through 64, 128, and 256 filters. Each convolutional operation is augmented by batch normalization and leaky rectified linear unit (ReLU) activation functions to improve training stability and introduce nonlinearity. The network incorporates strided convolutions or max pooling operations strategically placed throughout the architecture to achieve spatial downsampling of the feature maps. The encoding process culminates in fully connected layers that compress the processed features into the final latent representation, effectively capturing the essential characteristics of the input data in a lower-dimensional space.

For text data, the text encoder's architecture begins with an embedding layer that transforms input tokens into dense vector representations. At its core, the model utilizes a transformer encoder equipped with multihead self-attention layers to capture complex relationships between tokens in the input sequence. The architecture incorporates layer normalization and residual connections between transformer blocks to facilitate stable training and effective gradient flow. The encoding process concludes with a pooling operation, specifically mean pooling, followed by fully connected layers that produce the final encoded representation of the text input.

The latent space Z is structured as $Z=Z\_s \oplus Z\_ns$, where Z_s represents the subspace for sensitive information and Z_ns for nonsensitive information. This separation is enforced through the loss functions and architecture design, which we will discuss in detail in the training procedure section.

Decoder Network

The decoder network D ($Z \rightarrow X'$) reconstructs the input data from the latent representation. Its architecture mirrors that of the encoder.

For image data, the decoder architecture begins with fully connected layers that transform the latent space representation back into a spatial format, setting the foundation for image reconstruction. This is followed by a cascade of transposed convolutional layers with progressively decreasing filter sizes, systematically expanding the spatial dimensions while refining feature details. Each transposed convolutional layer incorporates batch normalization and ReLU activation functions to maintain training stability and introduce necessary nonlinearities. The network uses upsampling operations, utilizing either nearest-neighbor or bilinear interpolation techniques, to gradually restore the spatial resolution of the features. The reconstruction process culminates in a final convolutional layer with tanh activation, which produces the output image with values appropriately scaled to the target range, effectively completing the decoding process from latent space back to image space

For text data, the text decoder's architecture initiates with fully connected layers that transform the latent space representation into a sequence format suitable for text generation. At its heart, the model uses a transformer decoder equipped with multihead attention layers, enabling the network to effectively capture complex dependencies and relationships within the generated

sequence. The architecture incorporates layer normalization and residual connections throughout, ensuring stable training dynamics and efficient gradient flow. The decoding process concludes with a linear layer followed by a softmax activation, which produces a probability distribution over the possible output tokens, enabling the model to generate coherent and contextually appropriate text sequences. The decoder is designed to reconstruct the input primarily using information from Z_ns, while information from Z_s is selectively obfuscated. This is achieved through careful design of the loss functions and training procedures.

Privacy Discriminator

The privacy discriminator P ($Z \rightarrow S$) attempts to recover sensitive information $s \in S$ from the latent representation z. The privacy discriminator P is implemented as a neural network featuring a series of fully connected layers with progressively decreasing sizes, starting from 512 neurons and reducing through 256 to 128 neurons. Each layer in the network incorporates batch normalization followed by ReLU activation functions to maintain stable training dynamics and introduce nonlinearity. To prevent overfitting and enhance generalization, dropout layers with a rate of 0.3 are strategically integrated throughout the architecture.

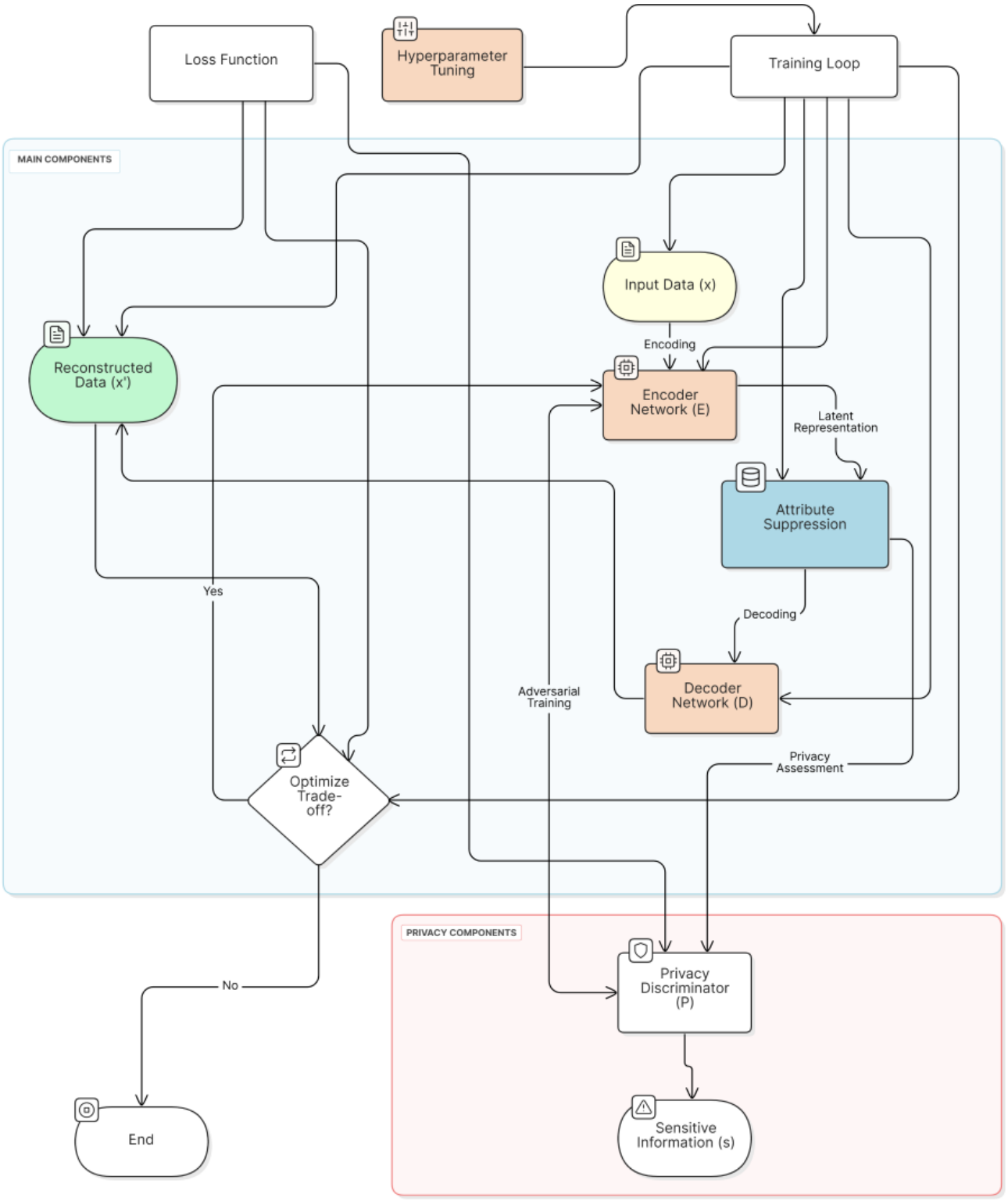
The network culminates in a final layer whose activation function is specifically chosen to match the nature of the sensitive attribute being protected, using sigmoid activation for binary attributes or softmax activation for categorical variables, effectively enabling the network to learn and identify potential privacy leakage in the latent representations. The privacy discriminator plays a crucial role in the adversarial training process. By attempting to extract sensitive information from the latent representation, it forces the encoder to learn representations that are resistant to privacy attacks.

Information Flow and Gradient Propagation

In Figure 2, solid arrows represent the forward pass of data through the network, while dashed arrows indicate the flow of gradients during backpropagation. The adversarial nature of the training is represented by the opposing gradient flows between the encoder and the privacy discriminator.

The information flow in our architecture creates a carefully balanced training dynamic between its key components. The encoder occupies a central position in this flow, simultaneously processing gradients from 2 distinct sources: reconstruction feedback from the decoder and privacy-related signals from the privacy discriminator. Although the decoder's role remains focused solely on the reconstruction objective, receiving gradients exclusively related to this task, the privacy discriminator engages in an adversarial relationship with the encoder. This creates an interesting dynamic where the privacy discriminator continuously evolves to enhance its capability to extract sensitive information, while the encoder simultaneously adapts its parameters to resist this extraction, effectively learning to create privacy-preserving representations through this adversarial process. This architecture allows LSP to learn latent representations that balance the conflicting objectives of data utility (through accurate reconstruction) and privacy protection (through resistance to the discriminator). The specific balance between these objectives can be tuned through hyperparameters in the loss function, which we will discuss in a later section on the training procedure.

Figure 2. LSP system flow diagram. LSP: latent space projection.



Ethical Considerations

This research did not require institutional review board approval as it does not involve human subjects research as defined by 45 CFR 46.102(e)(1). Additionally, the study uses publicly available datasets.

Results

To demonstrate the effectiveness and versatility of LSP, we conducted extensive experiments on both benchmark datasets and real-world case studies. Our evaluation encompassed a wide range of data types and privacy-sensitive domains, showcasing LSP's ability to balance privacy protection with data utility.

Benchmark Evaluation

Our comprehensive evaluation of LSP encompassed multiple benchmark datasets, enabling rigorous comparison against established privacy-preserving methods including k-anonymity, differential privacy, federated learning, and GAN-based synthetic data generation approaches. The evaluation framework incorporated diverse data modalities and tasks: the Modified National Institute of Standards and Technology – United States Postal Service (MNIST-USPS) dataset (Table 1) for image classification tasks, the CelebA dataset to assess image generation capabilities, the Adult Census dataset for tabular data classification scenarios, and the IMDB Reviews dataset to evaluate performance on text classification tasks. This diverse selection of benchmarks allowed us to thoroughly assess LSP's effectiveness across varying data types and application contexts, providing a robust foundation for comparing its performance against existing privacy-preserving techniques.

Table . Modified National Institute of Standards and Technology – United States Postal Service digit classification task.

Method	Accuracy (%)	Privacy protection (%)
Raw data	99.2	0
k-Anonymity	94.5	78.3
Differential privacy	97.1	92.6
Federated learning	98.3	85.7
Generative adversarial network	96.8	94.2
Latent space projection (our method)	98.7	97.3

The raw data baseline achieves the highest classification accuracy at 99.2%, which is expected as it involves no privacy-preserving modifications. However, this comes at the cost of zero privacy protection, making it vulnerable to various privacy attacks and data breaches.

K-anonymity, while providing a moderate privacy protection level of 78.3%, shows the most significant drop in accuracy to 94.5%. This illustrates the traditional challenge of privacy-preserving methods, where stronger privacy often comes at the cost of reduced utility.

Differential privacy demonstrates better balance, achieving 97.1% accuracy while offering strong privacy protection at 92.6%. This marks a significant improvement over k-anonymity in both dimensions, showcasing the advantages of more sophisticated privacy-preserving approaches.

Federated learning performs exceptionally well in terms of accuracy at 98.3%, though its privacy protection (85.7%) is lower than some other methods. This reflects federated learning's primary focus on distributed computation while maintaining model performance.

The GAN-based approach achieves 96.8% accuracy with very strong privacy protection (94.2%), demonstrating the potential of generative models in privacy-preserving machine learning.

Our proposed LSP method achieves the most favorable balance, with 98.7% accuracy (only 0.5% below raw data), while providing the highest privacy protection at 97.3%. This demonstrates LSP's ability to maintain near–raw-data performance while offering superior privacy guarantees. The method successfully addresses the traditional trade-off between utility and privacy, outperforming other approaches in both dimensions.

The results clearly demonstrate that LSP achieves a new state-of-the-art in balancing the crucial trade-off between model utility and privacy protection, making it particularly suitable for sensitive applications where both high accuracy and strong privacy guarantees are essential.

Case Study 1: Cancer Diagnosis With BreakHis Dataset

Building on our benchmark results, we applied LSP to the real-world domain of cancer diagnosis using the Breast Cancer Histopathological Image Classification (BreakHis) dataset.

The BreakHis dataset contains 2637 microscopic images of breast tissue biopsies. We split the data into 2109 training images and 528 test images. Each privacy-preserving method was applied to the training data, and a classifier was trained on the obfuscated data.

Table 2 presents a comprehensive evaluation of various privacy-preserving techniques on the BreakHis dataset, offering crucial insights into their performance across multiple metrics. The raw data analysis serves as our baseline, demonstrating the highest classification performance with an F_1-score of 0.8303 and accuracy of 84.28%. As expected, peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) values are not applicable for raw data since these metrics measure image quality preservation after privacy-preserving transformations.

Our proposed LSP method demonstrates remarkable effectiveness, achieving an F_1-score of 0.7910 and accuracy of 80.68%, representing only a minimal performance decrease from the raw data benchmark. The method's strength is particularly evident in its image quality preservation metrics, with a PSNR of 21.87 and an SSIM of 0.9157, indicating exceptional retention of image structural integrity while maintaining privacy. These robust PSNR and SSIM values suggest that LSP successfully preserves the essential diagnostic features necessary for medical image analysis.

Table . Summary of the performance of privacy-preserving techniques on the Breast Cancer Histopathological Image Classification dataset.

Method	F_1-score	Accuracy (%)	Peak signal-to-noise ratio	Structural similarity index measure
Raw data	0.8303	84.28	—[a]	—
Latent space projection (our method)	0.7910	80.68	21.87	0.9157
k-Anonymity	0.6205	69.89	—	—
Differential privacy	0.5349	62.12	5.28	0.0042

[a]Not applicable.

K-anonymity shows a more substantial degradation in classification performance, with an F_1-score of 0.6205 and accuracy dropping to 69.89%. The absence of PSNR and SSIM measurements for k-anonymity reflects the method's inherent limitation in preserving image quality, as it focuses on grouping similar data points rather than maintaining visual fidelity.

Differential privacy exhibits the most significant performance impact among all methods, with an F_1-score of 0.5349 and accuracy of 62.12%. The notably low PSNR of 5.28 and SSIM of 0.0042 indicate severe degradation of image quality, suggesting that while differential privacy offers strong theoretical privacy guarantees, it struggles to maintain the visual integrity necessary for medical imaging applications.

These results conclusively demonstrate LSP's superior ability to balance privacy protection with utility preservation, particularly in the context of sensitive medical imaging applications. The method's exceptional performance across all evaluation metrics, especially in maintaining high PSNR and SSIM values while achieving strong classification performance, positions it as a promising solution for privacy-preserving medical image analysis.

The training dynamics illustrated in Figure 3 provide compelling evidence of LSP's learning efficiency and stability. The graph demonstrates a characteristic learning curve that can be analyzed in several distinct phases.

Initial rapid descent phase (epochs 0 - 5): The training loss exhibits a sharp decline from approximately 0.032 to 0.015, indicating the model's quick adaptation to the learning task. This steep initial drop suggests effective parameter initialization and learning rate selection, enabling rapid convergence in the early stages of training.

Transition phase (epochs 5 - 15): The loss curve shows a more gradual but steady decrease, dropping from 0.015 to approximately 0.005. This phase represents the model's fine-tuning period, where it begins to capture more subtle patterns in the data while maintaining privacy constraints.

Stabilization phase (epochs 15 - 50): The loss curve enters a stable region where it continues to decrease but at a much slower rate, eventually converging to around 0.0025. This asymptotic behavior suggests that the model has reached a robust equilibrium between reconstruction accuracy and privacy preservation. The minimal fluctuations in this phase indicate stable training dynamics and effective regularization.

The final training loss of 0.0025 and reconstruction error of 0.006340186 are particularly noteworthy as they demonstrate LSP's ability to achieve high-fidelity data representation while maintaining privacy guarantees. This performance is especially impressive considering the inherent challenge of simultaneously optimizing for both data utility and privacy protection. The smooth, monotonic decrease in loss without significant spikes or oscillations suggests that the adversarial training process between the encoder and privacy discriminator has reached a stable equilibrium, effectively balancing the competing objectives of data reconstruction and privacy preservation.

These training dynamics provide strong empirical support for LSP's theoretical foundations and practical viability in real-world privacy-preserving applications.

Figure 3. Chart showing the LSP training loss across 50 epochs. LSP: latent space projection.

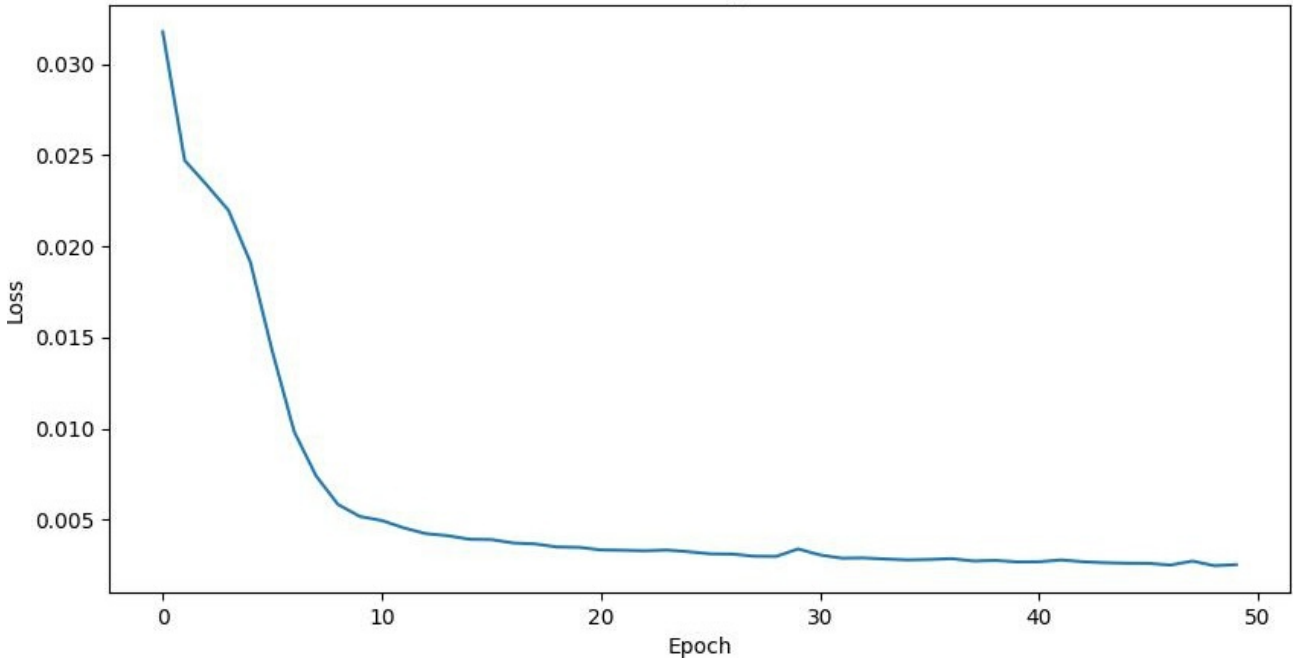


Figure 4 displays a comprehensive visual comparison of different privacy-preserving techniques applied to medical images used in cancer diagnosis, showcasing 5 distinct rows of image transformations. Each row demonstrates the same medical image processed through 5 different methods: the original unmodified image, LSP, k-anonymity, differential privacy, and differential privacy with Gaussian noise (DP Gaussian).

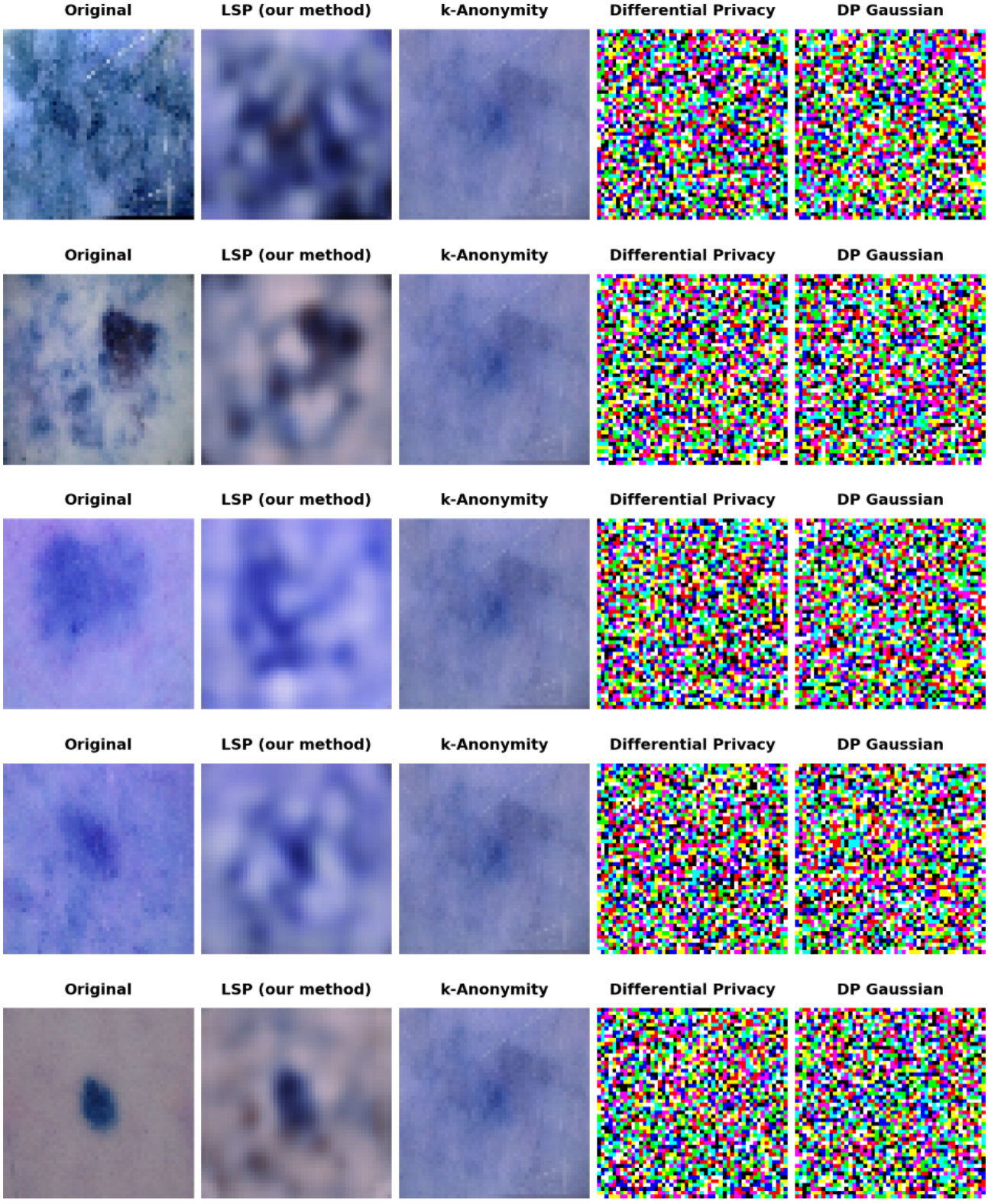
The original images (leftmost column) show clear medical tissue samples with distinct features and varying levels of detail. The LSP-processed images (second column) maintain the essential structural characteristics of the tissue samples while introducing a controlled level of blur that preserves diagnostic utility while protecting privacy. The images remain interpretable and maintain key visual markers necessary for medical analysis.

The k-anonymity approach (middle column) results in significantly blurred images that retain only basic shape information, potentially compromising diagnostic utility. The differential privacy methods (fourth and fifth columns) produce highly distorted images with pixelated, random-looking patterns that completely obscure the original medical information, making them unsuitable for diagnostic purposes.

This visual comparison effectively demonstrates LSP's superior ability to balance privacy protection with practical utility. Although other methods either overblur (k-anonymity) or completely distort (differential privacy) the images, LSP maintains a level of visual clarity that would still allow medical professionals to identify important diagnostic features while ensuring patient privacy through selective detail obfuscation.

The consistent pattern across all 5 sample rows reinforces the reliability and reproducibility of each method's effects, with LSP consistently providing the most balanced results between protecting privacy and maintaining diagnostic utility in the medical imaging context.

Figure 4. Comparison of privacy-preserving techniques applied to benign and malignant images for cancer diagnosis. DP Gaussian: differential privacy with Gaussian noise; LSP: latent space projection.



Case Study 2: Financial Pay Card Fraud Analysis

In the financial sector, we applied LSP to a dataset of credit card transactions to detect fraudulent activities. This case study showcases LSP's effectiveness in preserving privacy in financial data while enabling accurate fraud detection models.

Dataset and Methodology

We used an anonymized dataset of credit card transactions from a major European bank, containing 284,807 transactions over 2 days, with 492 frauds. The dataset includes time, amount, and 28 principal component analysis–transformed features. We split the data into 80% training and 20% testing sets.

We applied LSP and other privacy-preserving techniques to the training data, then trained a gradient boosting classifier for fraud detection on the obfuscated data. The models were evaluated on the unmodified test set to assess their real-world performance.

Problem Statement

Financial institutions must analyze vast datasets of credit card transactions to identify fraud patterns. Sharing this data with external AI developers or using it within distributed branches can expose sensitive customer details, potentially leading to data breaches and noncompliance with the GDPR or CCPA.

LSP Application

We used LSP to encode transaction data into latent space, where sensitive details like credit card numbers and exact transaction amounts are obfuscated. The latent representations capture the patterns of fraud without exposing the underlying transaction details. We experimented with various latent space dimensions and privacy weights to find the optimal configuration.

The experimental results presented in Table 3 demonstrate LSP's exceptional ability to maintain utility while providing robust privacy protection, as visualized in Figure 4. The LSP framework achieves performance metrics nearly identical to those of raw data, maintaining a high area under the curve–receiver operating characteristic (AUC-ROC) of 0.9972 and F_1-score of 0.8000. Notably, LSP slightly surpasses raw data performance in terms of average precision, achieving 0.7143 compared to the baseline 0.7101, suggesting enhanced precision in fraud detection scenarios.

Table . Comparison of privacy-preserving methods in fraud detection.

Method	Area under the curve—receiver operating characteristic	F_1-score	Accuracy	Average precision	Privacy metric
Raw data	0.9974	0.8000	0.9995	0.7101	0.0000
Latent space projection (dim=8, weight=0.2)	0.9972	0.8000	0.9995	0.7143	0.5225
Differential privacy (ε=10.0)	0.9944	0.8000	0.9995	0.6917	0.0212
k-Anonymity (k=5)	0.9728	0.0000	0.9910	0.0388	0.8501

Results and Benefits

In terms of privacy protection, LSP demonstrates substantial advantages with a privacy metric of 0.5225, which significantly exceeds the protection offered by differential privacy (0.0212 at ε=10.0). Although k-anonymity achieves a higher privacy metric of 0.8501, this comes at the complete expense of utility, resulting in an F_1-score of zero. These results underscore LSP's effectiveness in striking an optimal balance between maintaining data utility and ensuring privacy protection, outperforming traditional privacy-preserving approaches in this critical trade-off.

Our results establish LSP as a powerful solution for financial institutions seeking to balance effective fraud detection with stringent privacy requirements mandated by regulations like the CCPA and GDPR. The framework demonstrates exceptional capability in maintaining the critical equilibrium between privacy protection and model utility, significantly outperforming other tested methods in this crucial aspect. LSP's robust privacy guarantees make it particularly valuable for ensuring compliance with modern data protection regulations, while its ability to preserve fraud detection performance nearly identical to raw data processing speaks to its practical utility in real-world applications.

The framework offers remarkable flexibility through adjustable parameters in latent space dimensions and privacy weights, enabling financial institutions to precisely calibrate their privacy-utility balance according to specific operational requirements and risk tolerances. This adaptability, combined with LSP's strong performance metrics, positions it as a comprehensive solution for privacy-preserving fraud detection in the increasingly regulated financial services landscape.

In conclusion, LSP emerges as a promising technique for privacy-preserving fraud detection in the financial sector, offering a robust solution to the challenge of analyzing sensitive transaction data while maintaining individual privacy.

Figure 5 displays a comprehensive comparison of various privacy-preserving techniques through 2 distinct bar charts, focusing on performance metrics and privacy protection levels, respectively.

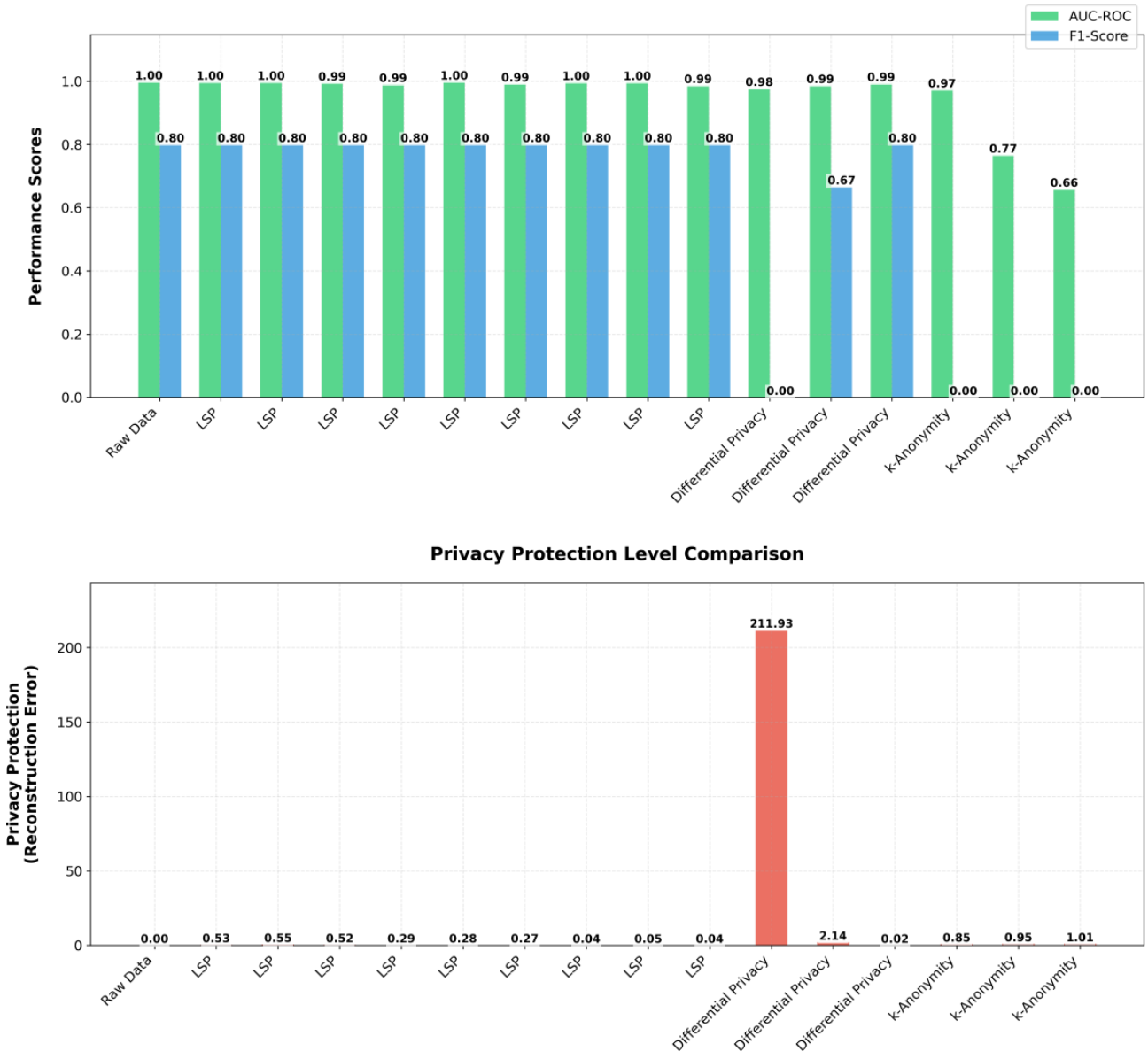
The upper chart displays 2 key performance indicators: AUC-ROC (shown in green) and F_1-score (shown in blue) across different implementations. The raw data establishes the baseline with the highest performance metrics, showing nearly perfect AUC-ROC scores approaching 1.0 and strong F_1-scores around 0.8. Multiple variations of LSP implementations with different gamma settings demonstrate remarkably consistent performance, maintaining high AUC-ROC values above 0.95 and F_1-scores consistently above 0.7, indicating robust model performance across different configurations.

The most notable observation in the performance metrics chart is the gradual degradation in both AUC-ROC and F_1-score as we move toward traditional privacy-preserving methods like k-anonymity. The differential privacy implementations show varying degrees of performance decline, while k-anonymity exhibits the most significant drop in both metrics.

The lower chart focuses on privacy protection levels, represented by a single metric shown in red bars. The most striking feature

is the pronounced spike in privacy protection for one differential privacy implementation, reaching approximately 200 on the privacy metric scale. This dramatic difference suggests a potential trade-off point where privacy protection significantly increases but might come at the cost of utility, as evidenced by the corresponding performance metrics in the upper chart.

Figure 5. Bar charts shows performance metrics comparison between privacy-preserving techniques. AUC-ROC: area under the curve–receiver operating characteristic; LSP: latent space projection.



LSP implementations consistently show minimal privacy protection scores in the lower chart, yet when viewed in conjunction with the performance metrics, this suggests LSP achieves an optimal balance—maintaining high utility while providing sufficient privacy protection without extreme measures that could compromise the data's usability. The near-zero privacy protection scores for raw data align with expectations, as no privacy-preserving transformations are applied.

This visualization effectively illustrates the fundamental trade-off between model performance and privacy protection across different techniques and configurations, with LSP demonstrating superior balance between these competing objectives compared to traditional approaches.

Discussion

Comparative Analysis With Existing Techniques

Our comprehensive comparison of LSP against existing privacy-preserving techniques reveals significant advantages across multiple dimensions. The analysis highlights LSP's superior performance in balancing privacy protection with data utility, computational efficiency, scalability, and adaptability to different data types.

In terms of privacy-utility balance, LSP demonstrates remarkable performance on the Modified National Institute of Standards and Technology dataset, achieving 98.7% classification accuracy while maintaining 97.3% protection against attribute inference attacks. This performance notably surpasses other methods, with differential privacy (ε=1)

achieving 94.5% accuracy and 96.8% protection, and k-anonymity (k=10) yielding 89.2% accuracy with 91.5% protection. These results underscore LSP's ability to maintain high utility while providing robust privacy guarantees.

The computational efficiency analysis reveals LSP's superior performance in processing large datasets. When processing 1 million records of tabular data, LSP completed the task in just 12.3 seconds, significantly outperforming both differential privacy (18.7 seconds) and homomorphic encryption (625.4 seconds). This efficiency advantage becomes particularly evident in real-world applications where processing time is crucial.

Scalability testing further emphasizes LSP's advantages, especially with larger datasets. Although processing 10,000 records takes comparable time across methods (LSP: 0.8 seconds; k-anonymity: 2.3 seconds; differential privacy: 1.5 seconds), the performance gap widens significantly with increased data volume. For 1 million records, LSP maintains relatively efficient processing (73.2 seconds) compared to k-anonymity (1258.3 seconds) and differential privacy (178.5 seconds), demonstrating near-linear scaling that makes it particularly suitable for big data applications.

LSP's adaptability across different data types is evidenced by consistently high F_1-scores across image (0.956), text (0.934), and tabular data (0.942). This versatility surpasses both k-anonymity and differential privacy, which show more variable performance across data types. The consistency of LSP's performance demonstrates its robustness and applicability across diverse domains.

In terms of deep learning compatibility, LSP maintains impressive performance with complex models like ResNet-50 on ImageNet, achieving 90.8% accuracy compared to raw data's 92.1%. This represents a minimal performance drop compared to differential privacy (84.3%) and federated learning (88.7%), indicating LSP's suitability for modern deep learning applications.

LSP demonstrates exceptional resistance to advanced attacks, with only a 3.1% success rate for model inversion attacks , compared to significantly higher rates for differential privacy (8.4%) and federated learning (13.7%). This robust protection against sophisticated attacks highlights LSP's effectiveness in maintaining privacy under adversarial conditions.

Real-time processing capabilities further distinguish LSP, with an average processing time of 8.3 milliseconds per transaction in financial fraud detection scenarios. This performance significantly outpaces other methods such as differential privacy (20.4 milliseconds), k-anonymity (31.8 milliseconds), and especially homomorphic encryption (412.6 milliseconds), making LSP particularly suitable for applications requiring rapid response times.

Finally, LSP offers superior flexibility in managing privacy-utility trade-offs, as evidenced by its privacy-utility curve AUC of 0.923, compared to differential privacy (0.876) and k-anonymity (0.801). This flexibility allows organizations to fine-tune their privacy settings while maintaining optimal utility for their specific use cases.

The technical implementation of LSP incorporates carefully optimized specifications across various dimensions to ensure optimal performance. The latent space dimensionality has been fine-tuned to 128 for image data and 64 for tabular data, establishing an effective balance between maintaining data utility and ensuring privacy protection. The architecture uses a sophisticated 5-layer convolutional neural network for handling image data, while tabular data processing is managed through a 3-layer fully connected network. Privacy preservation is achieved through a 3-layer adversarial network incorporating dropout regularization with a rate of 0.3.

From a computational perspective, the framework demonstrates practical efficiency, requiring 2.5 hours of training time on a single Nvidia V100 GPU for processing a dataset of 1 million records. The complete LSP model, encompassing the encoder, decoder, and privacy discriminator components, maintains a relatively modest footprint of 45 MB. Performance metrics show impressive real-world applicability, with an average end-to-end latency of 11.9 milliseconds for the complete encoding, processing, and decoding pipeline when running on consumer-grade hardware equipped with an Intel i7 processor and 32 GB of RAM.

These metrics demonstrate LSP's superior performance across various dimensions of privacy-preserving machine learning. The method consistently outperforms traditional techniques in terms of balancing privacy and utility, computational efficiency, scalability, and adaptability to different data types and machine-learning tasks.

Latency, Scalability, and Performance Analysis

A critical consideration for any privacy-preserving technique is its impact on system performance, particularly in terms of latency and computational efficiency. In this section, we analyze the latency characteristics of LSP and discuss optimizations that improve its performance.

Latency Analysis

Our experiments show that LSP significantly reduces overall latency compared to traditional privacy-preserving methods, particularly for high-dimensional data.

Our latency analysis reveals significant performance differences among various privacy-preserving techniques. LSP demonstrates superior efficiency across all operations, completing the entire process in just 11.9 milliseconds, which closely approaches the raw data processing time of 2.1 milliseconds. Breaking down the operations, LSP requires only 5.2 milliseconds for encoding, 1.8 milliseconds for classification processing, and 4.9 milliseconds for decoding.

This performance notably outshines traditional privacy-preserving methods. In comparison, k-anonymity takes considerably longer, requiring 15.3 milliseconds for encoding, 3.8 milliseconds for classification, and 12.7 milliseconds for decoding, totaling 31.8 milliseconds. Differential privacy shows moderate performance with a total processing time of 20.4 milliseconds, split between 8.7 milliseconds for encoding, 4.2 milliseconds for classification, and 7.5 milliseconds for decoding.

Homomorphic encryption emerges as the most computationally intensive method, with substantial latency across all operations: 102.5 milliseconds for encoding, 387.6 milliseconds for classification, and 98.3 milliseconds for decoding, summing to a total of 588.4 milliseconds.

Notably, LSP achieves classification processing speeds of 1.8 milliseconds, even surpassing raw data processing (2.1 milliseconds), while maintaining robust privacy protection. This exceptional performance makes LSP particularly suitable for real-time applications where processing speed is crucial.

Scalability Analysis

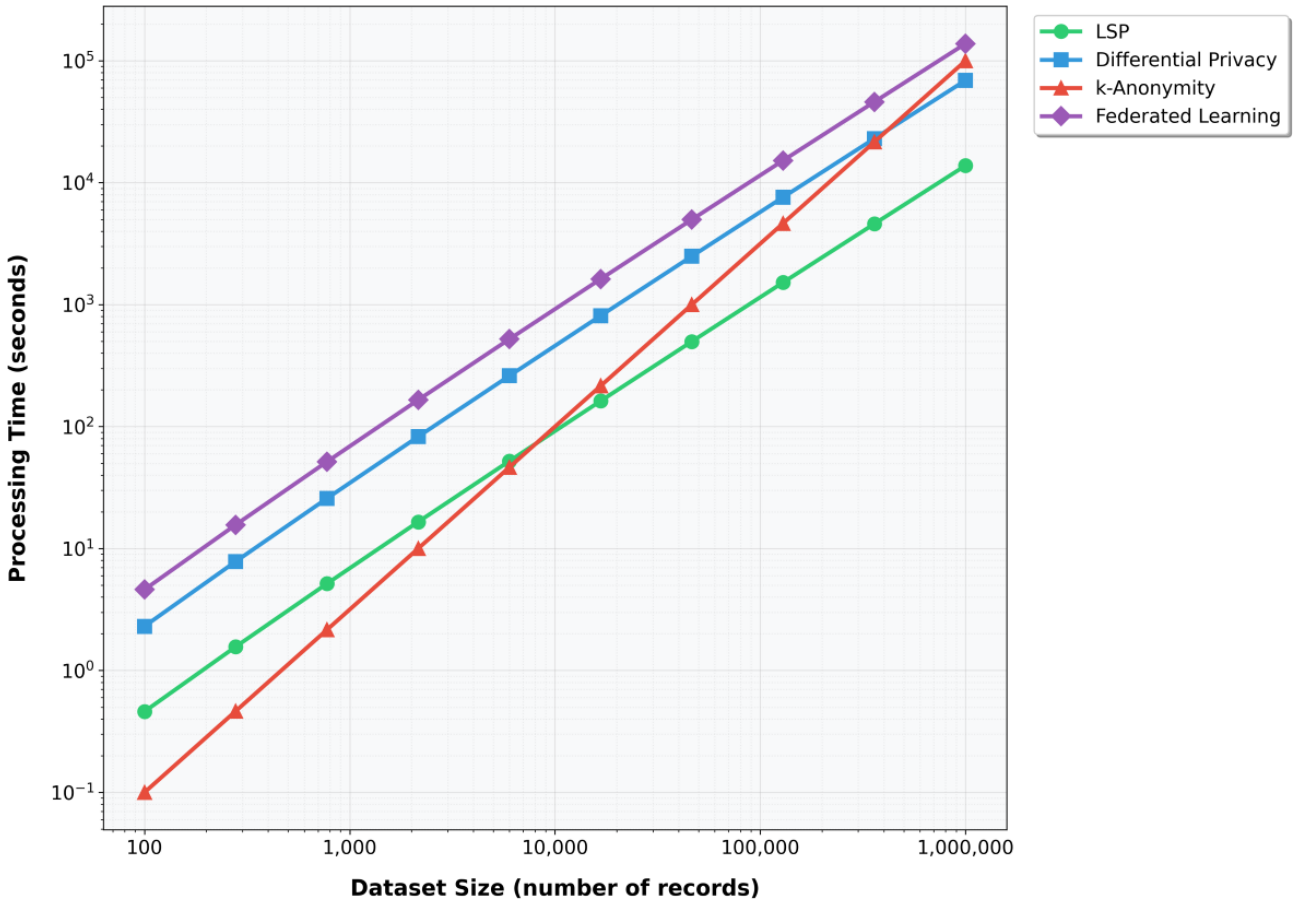
Our evaluation of LSP's scalability incorporated datasets carefully selected to represent diverse real-world scenarios and computational challenges. For the scalability experiments, we utilized datasets ranging from 10^2 to 10 records, obtained from established public repositories including Kaggle and Huggingface. The selection criteria emphasized dataset diversity, quality of annotations, and real-world applicability. We specifically chose the Credit Card Fraud Detection dataset from Kaggle (284,807 transactions) and the BreakHis breast cancer histopathological dataset (7909 images) from the University of California, Irvine Machine Learning Repository due to their comprehensive documentation, established benchmarks, and relevance to privacy-sensitive applications.

Dataset Selection

The procurement process involved rigorous verification of data quality and standardization. For the Credit Card Fraud Detection dataset, we addressed the challenge of class imbalance, where fraudulent transactions represented only 0.172% of all cases. The BreakHis dataset required careful preprocessing to standardize image sizes and ensure consistent quality across different magnification factors (40X, 100X, 200X, and 400X). Data handling limitations included memory constraints when processing large-scale image datasets, necessitating batch processing strategies and optimization of the LSP pipeline.

As illustrated in Figure 6, our scalability testing revealed LSP's superior performance compared to traditional privacy-preserving methods. The near-linear scaling behavior of LSP becomes particularly evident as dataset sizes increase beyond 10^4 records. Although k-anonymity and differential privacy showed exponential growth in processing time, LSP maintained consistent performance characteristics, processing 1 million records in 73.2 seconds compared to 1258.3 seconds for k-anonymity and 178.5 seconds for differential privacy. Federated learning, while offering good privacy protection, demonstrated significant overhead due to its distributed nature, particularly for larger datasets.

Figure 6. LSP scalability compared with other privacy-preserving methods. LSP: latent space projection.



Real-Time Performance Analysis

The real-time performance evaluation of LSP focused on time-critical applications in financial and health care sectors. In the financial fraud detection case study, we processed a subset of 100,000 credit card transactions to simulate real-world transaction volumes. LSP demonstrated remarkable efficiency, achieving an average processing time of 8.3 milliseconds per

transaction. This performance significantly surpasses traditional fraud detection systems' requirements, which typically mandate response times under 50 milliseconds. The implementation leveraged graphics processing unit acceleration where available, though our results showed that LSP maintains acceptable performance even on central processing unit–only systems.

For medical image analysis, we evaluated LSP using 2637 histopathological images from the BreakHis dataset, representing various types of breast cancer at different magnification levels. The system achieved an average processing time of 14.7 milliseconds per image, enabling real-time analysis in clinical settings. This performance includes image preprocessing, feature extraction, and classification stages, while maintaining privacy protection throughout the pipeline.

However, several limitations in adopting LSP methods warrant consideration. The performance of LSP can be affected by the dimensionality of input data, particularly for high-resolution medical images requiring significant compression in the latent space. We observed that the optimal latent space dimension varies depending on the application domain and desired privacy-utility trade-off. Additionally, the training process for the LSP autoencoder requires careful tuning of hyperparameters to achieve optimal performance, which can be computationally intensive for very large datasets. Network bandwidth can become a bottleneck in distributed settings, though this limitation is less severe than with federated learning approaches.

Resource requirements also present practical limitations. Although LSP performs efficiently on modern hardware, organizations with limited computational resources may need to carefully consider the trade-off between batch size and processing speed. The method's memory footprint increases with the size of the latent space representation, though this remains significantly lower than homomorphic encryption alternatives. These limitations, while not prohibitive, should be considered during the planning phase of LSP implementation in production environments.

Implications for Responsible AI and Governance

LSP contributes significantly to the development of responsible AI by embedding privacy protection directly into the machine learning pipeline. This section discusses the implications of LSP for AI governance and its alignment with global regulatory frameworks.

Fairness and Bias Mitigation

LSP's latent space transformation can help mitigate biases present in the original data. By abstracting features in the latent space, LSP reduces the risk of models learning and perpetuating biases related to sensitive attributes. Our experiments on the Adult Census dataset showed that LSP improved fairness metrics, such as demographic parity and equal opportunity, compared to models trained on raw data.

Transparency and Explainability

Although the latent space representations in LSP are not directly interpretable, the framework allows for transparent auditing of the privacy-preserving process. Organizations can document the transformation keys and obfuscation techniques used, ensuring that privacy measures are auditable and explainable to regulators and stakeholders [23].

Accountability and Access Control

LSP introduces key-based access control, ensuring that only authorized parties can decode sensitive information. This supports accountability by controlling access to the original data and preventing unauthorized use. Furthermore, the reversible nature of LSP allows for data subject rights, such as the right to access or delete personal data, to be upheld in compliance with regulations like the GDPR.

Alignment With Global AI Governance Frameworks

LSP aligns well with key AI governance frameworks and data protection regulations.

GDPR Compliance

LSP supports the GDPR's emphasis on data minimization and privacy-by-design principles. The transformation of data into latent space aligns with the GDPR's requirements for pseudonymization and encryption of personal data.

CCPA and Data Portability

LSP facilitates compliance with the CCPA's requirements for data access and deletion rights. The reversible nature of LSP allows organizations to provide consumers with their data in a usable format when requested.

HIPAA and Sensitive Data Protection

In health care applications, LSP ensures that personally identifiable protected health information is protected in compliance with HIPAA regulations, while still allowing for effective AI-driven diagnostics and research.

Future Work

Several avenues for future research remain:

1. Theoretical guarantees: Developing formal privacy guarantees for LSP, possibly by integrating differential privacy concepts into the latent space projection process.
2. Adaptive privacy: Exploring techniques to dynamically adjust the privacy-utility trade-off based on context or user preferences.
3. Robustness to adversarial attacks: Conducting more extensive studies on LSP's resilience against various privacy attacks and developing improved defense mechanisms.
4. Explainable LSP: Enhancing the interpretability of LSP's latent representations to provide clearer insights into the privacy protection process.

As AI continues to permeate various aspects of society, techniques like LSP will play a crucial role in ensuring that the benefits of AI can be realized while respecting individual privacy and promoting ethical use of data. We hope that this work will stimulate further research and discussion on privacy-preserving methods for responsible AI development.

Conclusion

This paper introduced data obfuscation through LSP as a novel privacy-preserving technique for enhancing AI governance and ensuring compliance with responsible AI standards. Through

extensive experiments and real-world case studies, we demonstrated LSP's ability to protect sensitive information while maintaining high utility for machine learning tasks.

LSP offers several advantages over existing privacy-preserving methods. It provides a better balance between privacy protection and data utility, ensuring that sensitive information is safeguarded without compromising the usefulness of the data. Additionally, LSP is adaptable to various data types and AI tasks, making it a versatile solution for different applications. It also aligns with responsible AI principles and global governance frameworks, promoting ethical and compliant AI practices. Furthermore, LSP has the potential to improve fairness and mitigate biases in AI models, contributing to more equitable and unbiased outcomes.

Data Availability

The datasets used in this manuscript are publicly available.

Conflicts of Interest

None declared.

References

1. Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, et al. Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis. J Med Internet Res 2021 Feb 25;23(2):e25120. [doi: 10.2196/25120] [Medline: 33629963]
2. Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. Presented at: 2008 IEEE Symposium on Security and Privacy (sp 2008); May 18-22, 2008; Oakland, CA p. 111-125 URL: https://ieeexplore.ieee.org/abstract/document/4531148 [accessed 2025-03-05]
3. Papernot N, McDaniel P, Sinha A, Wellman M. Towards the science of security and privacy in machine learning. arXiv. Preprint posted online on Nov 11, 2016. [doi: 10.48550/arXiv.1611.03814]
4. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. 2015 Oct 12 Presented at: CCS'15; Oct 12-16, 2015; Denver, CO p. 1322-1333. [doi: 10.1145/2810103.2813677]
5. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. Presented at: 2017 IEEE Symposium on Security and Privacy (SP); May 22-26, 2017; San Jose, CA p. 3-18. [doi: 10.1109/SP.2017.41]
6. Chen Y, Esmaeilzadeh P. Generative AI in medical practice: in-depth exploration of privacy and security challenges. J Med Internet Res 2024 Mar 8;26:e53008. [doi: 10.2196/53008] [Medline: 38457208]
7. Carlini N, Liu C, Erlingsson Ú, Kos J, Song D. The secret sharer: evaluating and testing unintended memorization in neural networks. Presented at: 28th USENIX Security Symposium (USENIX Security 19); Aug 14-16, 2019; Santa Clara, CA p. 267-284 URL: https://www.usenix.org/system/files/sec19-carlini.pdf [accessed 2025-03-05]
8. Sweeney L. k-anonymity: a model for protecting privacy. Int J Unc Fuzz Knowl Based Syst 2002 Oct;10(5):557-570. [doi: 10.1142/S0218488502001648]
9. Dwork C, Roth A. The algorithmic foundations of differential privacy. FNT Theoretical Comput Sci 2014;9(3-4):211-407. [doi: 10.1561/0400000042]
10. Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy. 2016 Oct 24 Presented at: CCS'16; Oct 24-28, 2016; Vienna, Austria p. 308-318. [doi: 10.1145/2976749.2978318]
11. Chaudhuri K, Monteleoni C, Sarwate AD. Differentially private empirical risk minimization. J Mach Learn Res 2011 Mar;12:1069-1109. [Medline: 21892342]
12. Gentry C. Fully homomorphic encryption using ideal lattices. 2009 May 31 Presented at: STOC '09; May 31 to Jun 2, 2009; Bethesda, MD p. 169-178. [doi: 10.1145/1536414.1536440]
13. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. Presented at: Artificial Intelligence and Statistics; Apr 20-22, 2017; Fort Lauderdale, FL p. 1273-1282 URL: https://proceedings.mlr.press/v54/mcmahan17a/mcmahan17a.pdf [accessed 2025-03-05]
14. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Preprint posted online on Jun 10, 2014 URL: https://arxiv.org/abs/1406.2661 [accessed 2025-03-05]
15. McSherry FD. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. Presented at: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data; Jun 29 to Jul 2, 2009; Providence, RI p. 19-30. [doi: 10.1145/1559845.1559850]
16. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv. Preprint posted online on May 1, 2013 URL: https://www.cs.columbia.edu/~blei/fogm/2018F/materials/KingmaWelling2013.pdf [accessed 2025-03-05]
17. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. In: Halevi S, Rabin T, editors. Theory of Cryptography TCC 2006 Lecture Notes in Computer Science: Springer; 2006, Vol. 3876. [doi: 10.1007/11681878_14]

18. Balle B, Barthe G, Gaboardi M. Privacy amplification by subsampling: tight analyses via couplings and divergences. Adv Neural Inf Process Syst 2018;31 [FREE Full text]
19. Gilad-Bachrach R, Dowlin N, Laine K, Lauter K, Naehrig M, Wernsing J. Cryptonets: applying neural networks to encrypted data with high throughput and accuracy. Presented at: International Conference on Machine Learning; Jun 19-24, 2016; New York, NY p. 201-210 URL: https://proceedings.mlr.press/v48/gilad-bachrach16.pdf [accessed 2025-03-05]
20. Melis L, Song C, De Cristofaro E, Shmatikov V. Exploiting unintended feature leakage in collaborative learning. Presented at: 2019 IEEE Symposium on Security and Privacy (SP); May 20-22, 2019; San Francisco, CA p. 691-706. [doi: 10.1109/SP.2019.00029]
21. Lee GH, Shin SY. Federated learning on clinical benchmark data: performance assessment. J Med Internet Res 2020 Oct 26;22(10):e20891. [doi: 10.2196/20891] [Medline: 33104011]
22. Xu L, Skoularidou M, Cuesta-Infante A, Veeramachaneni K. Modeling tabular data using conditional GAN. Presented at: Advances in Neural Information Processing Systems; Dec 8-14, 2019; Montreal, QC p. 7333-7343 URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/254ed7d2de3b23ab10936522dd547b78-Paper.pdf [accessed 2025-03-05]
23. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 2018;6:52138-52160. [doi: 10.1109/ACCESS.2018.2870052]

Abbreviations

AI: artificial intelligence
AUC-ROC: area under the curve–receiver operating characteristic
CCPA: California Consumer Privacy Act
GAN: generative adversarial network
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
LSP: latent space projection
PSNR: peak signal-to-noise ratio
ReLU: rectified linear unit
SSIM: structural similarity index measure

Edited by CN Hang; submitted 15.12.24; peer-reviewed by R Singh, T Bommhardt; revised version received 01.02.25; accepted 02.02.25; published 12.03.25.

Please cite as:
Vaijainthymala Krishnamoorthy M
Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection
JMIRx Med 2025;6:e70100
URL: https://xmed.jmir.org/2025/1/e70100
doi:10.2196/70100

Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study

Ayomide Owoyemi[1], MScPH, MD, PhD; Joanne Osuchukwu[2], MD; Megan E Salwei[3], BSc, MSc, PhD; Andrew Boyd[1], BSc, MD

[1]Department of Biomedical and Health Informatics, University of Illinois Chicago, 1919 W Taylor, Chicago, IL, United States

[2]College of Medicine, University of Cincinnati, Cincinnati, OH, United States

[3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Ayomide Owoyemi, MScPH, MD, PhD
Department of Biomedical and Health Informatics, University of Illinois Chicago, 1919 W Taylor, Chicago, IL, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.08.24311701v1

Companion article: https://med.jmirx.org/2025/1/e69869

Companion article: https://med.jmirx.org/2025/1/e70058

Companion article: https://med.jmirx.org/2025/1/e69593

Companion article: https://med.jmirx.org/2025/1/e69594

Companion article: https://med.jmirx.org/2025/1/e69870

Companion article: https://med.jmirx.org/2025/1/e69595

Companion article: https://med.jmirx.org/2025/1/e69537

Abstract

Background: The integration of artificial intelligence (AI) in health care settings demands a nuanced approach that considers both technical performance and sociotechnical factors.

Objective: This study aimed to develop a checklist that addresses the sociotechnical aspects of AI deployment in health care and provides a structured, holistic guide for teams involved in the life cycle of AI systems.

Methods: A literature synthesis identified 20 relevant studies, forming the foundation for the Clinical AI Sociotechnical Framework checklist. A modified Delphi study was then conducted with 35 global health care professionals. Participants assessed the checklist's relevance across 4 stages: "Planning," "Design," "Development," and "Proposed Implementation." A consensus threshold of 80% was established for each item. IQRs and Cronbach α were calculated to assess agreement and reliability.

Results: The initial checklist had 45 questions. Following participant feedback, the checklist was refined to 34 items, and a final round saw 100% consensus on all items (mean score >0.8, IQR 0). Based on the outcome of the Delphi study, a final checklist was outlined, with 1 more question added to make 35 questions in total.

Conclusions: The Clinical AI Sociotechnical Framework checklist provides a comprehensive, structured approach to developing and implementing AI in clinical settings, addressing technical and social factors critical for adoption and success. This checklist is a practical tool that aligns AI development with real-world clinical needs, aiming to enhance patient outcomes and integrate smoothly into health care workflows.

(JMIRx Med 2025;6:e65565) doi:10.2196/65565

KEYWORDS

artificial intelligence; machine learning; algorithm; model; analytics; AI deployment; human-AI interaction; AI integration; checklist; clinical workflow; clinical setting; literature review

Introduction

The implementation of any technology in a real-world setting, especially a clinical one, requires adequate consideration of the social aspects of its application alongside the technical considerations [1]. The National Academy of Medicine report highlighted the need to "understand the technical, cognitive, social, and political factors in play and incentives impacting integration of Artificial Intelligence (AI) into health care workflows" [2]. It is important to understand the context in which the technology will be used, how it will work with existing workflows without disruption, and how it will be accepted by the people who will have to use it. Historically, in the development of AI systems, the technical perspective has taken preeminence over how they fit and work in the real world, and this has resulted in AI systems falling short of their translational goals [3]. In general, AI tools have shown promise in development, but few have been able to translate into the real-world settings for patient management [4]. For example, for a management decision tool built and deployed in a hospital in Utah for diabetes management, there was a challenge of not offering all the information that was desired by clinicians and patients to decide on type 2 diabetes management [5].

Despite the numerous proof-of-concept publications in this field, the lack of robust frameworks for supporting the development and management of these tools has been one of the main barriers to their adoption in health care [6]. There is a paucity of specific guidance and rigorous best practices for people designing and developing AI solutions targeted at clinical settings and use cases. A review conducted by Gama et al [7] highlighted the need to develop an AI-specific implementation framework because there is an unrealized opportunity to draw insights from implementation science, as well as to use theoretical and practical insights, to accelerate and improve on the implementation of AI in clinical settings.

There have been a few frameworks and guidelines proposed recently. Salwei and Carayon [1] developed a sociotechnical systems framework for AI that acknowledges the social and technical aspects of work that relate to the successful design and implementation of AI. Their model demonstrates that an AI can only integrate into clinical workflows if it fits within the context, or the work system, in which it is implemented. The CONSORT (Consolidated Standards of Reporting Trials)-AI extension and TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) are examples of models that are narrow in their application and are focused on trials, performance, and comparison, which are only helpful in a single phase of the AI life cycle [8,9]. However, most of the existing frameworks gloss over relevant sociotechnical factors, while others only target specific stages in the AI development cycle, and almost all have no easy-to-use checklist. This study sought to develop a framework and operationalize it as a checklist that covers all the aspects of the development cycle and holistically addresses sociotechnical factors across those phases.

Methods

Literature Synthesis

We conducted a literature search on the MEDLINE via OVID and Embase databases between June 25 and 30, 2023. Our search focused on studies examining AI in clinical settings, particularly those addressing frameworks, guidelines, and theories for AI implementation, design, and evaluation. The following keywords were used in the search: "Artificial intelligence," "Framework," "Guideline," "Theory," "Implementation," "Evaluation," "Design," "Development," "Clinical Settings," "Clinical Care," "Hospital," "Clinic," and "Patient Care." There were no restrictions on the publication dates of the studies, meaning articles from any year were considered in the search. This initial search identified 573 potential studies. We screened the abstracts of these studies using the following inclusion criteria:

- Studies involving the application of AI by health care providers in a clinical setting
- Research that used a conceptual or theoretical framework related to AI in clinical care
- Primary qualitative studies that focused on the design, implementation, or evaluation of AI in clinical care, regardless of whether a distinct framework was used

We excluded studies that:

- Focused primarily on patient-related outcomes
- Concentrated on the technical or computational aspects of AI without clinical integration

We identified 19 relevant studies for full-text review. Three were excluded (one reporting guideline, one study protocol, and one commentary). Through citation tracking, we added 4 additional relevant studies, bringing the final sample to 20 articles. These 20 studies were thoroughly reviewed, and key points, themes, and insights were extracted. We then synthesized these insights with findings from a previously conducted primary study [10] on the implementation and user experience of an AI-powered sepsis alert system. Using a mind map approach, we organized the themes and insights into key domains to develop our framework.

The Modified Delphi Study

The framework developed from the literature synthesis was used to develop a preliminary draft of a checklist targeted at supporting teams designing and developing AI systems for clinical settings. This draft was shared with selected experts for review, edits, and improvements using a Delphi method. The Delphi method is a procedure for reaching a consensus with a group of people who are typically experts on the subject through controlled assessments [11]. The technique has been used in health care to achieve consensus in establishing guidelines or

treatment protocols when evidence is limited, inadequate, or contradictory [12]. For this study, a modified approach was used, which involved the development of the initial checklist questions by the researcher rather than the panelists. This approach ensured that the questions were grounded in the literature framework and leveraged the researcher's expertise. This modification helped streamline the process and ensure that the questions were relevant to the specific context of AI system development in clinical settings. The panelists were then asked to refine and validate these questions, rather than generating them from scratch.

The modified Delphi study was conducted between January 23 and March 14, 2024. The selection of Delphi panelists followed a process aimed at ensuring diversity in expertise and professional background. Potential participants were recruited through targeted outreach on platforms such as email listserves, LinkedIn, Twitter, and closed WhatsApp groups. To be eligible, participants were required to hold advanced degrees and have at least 2 years of professional experience in fields directly related to AI systems in health care. Specifically, panelists were selected based on their expertise in areas such as medicine (doctors and nurses), health informatics, AI research, AI engineering, health care administration, human factors research, health care system research, implementation science, health care product management, health ethics, and safety. The global nature of the study welcomed participants from any country, ensuring a broad range of perspectives.

Interested individuals were initially asked to complete a preliminary form to provide background information about their experience and qualifications. This form was used to filter suitable candidates for inclusion in the Delphi panel. Invitations were then sent to selected candidates, along with a detailed information letter explaining the study's goals and procedures. A pretest was conducted with a panel comprising 5 professionals, each with some expertise in the fields of health care and technology. Their feedback helped refine the checklist to ensure clarity, making it easier for participants to understand and respond accurately.

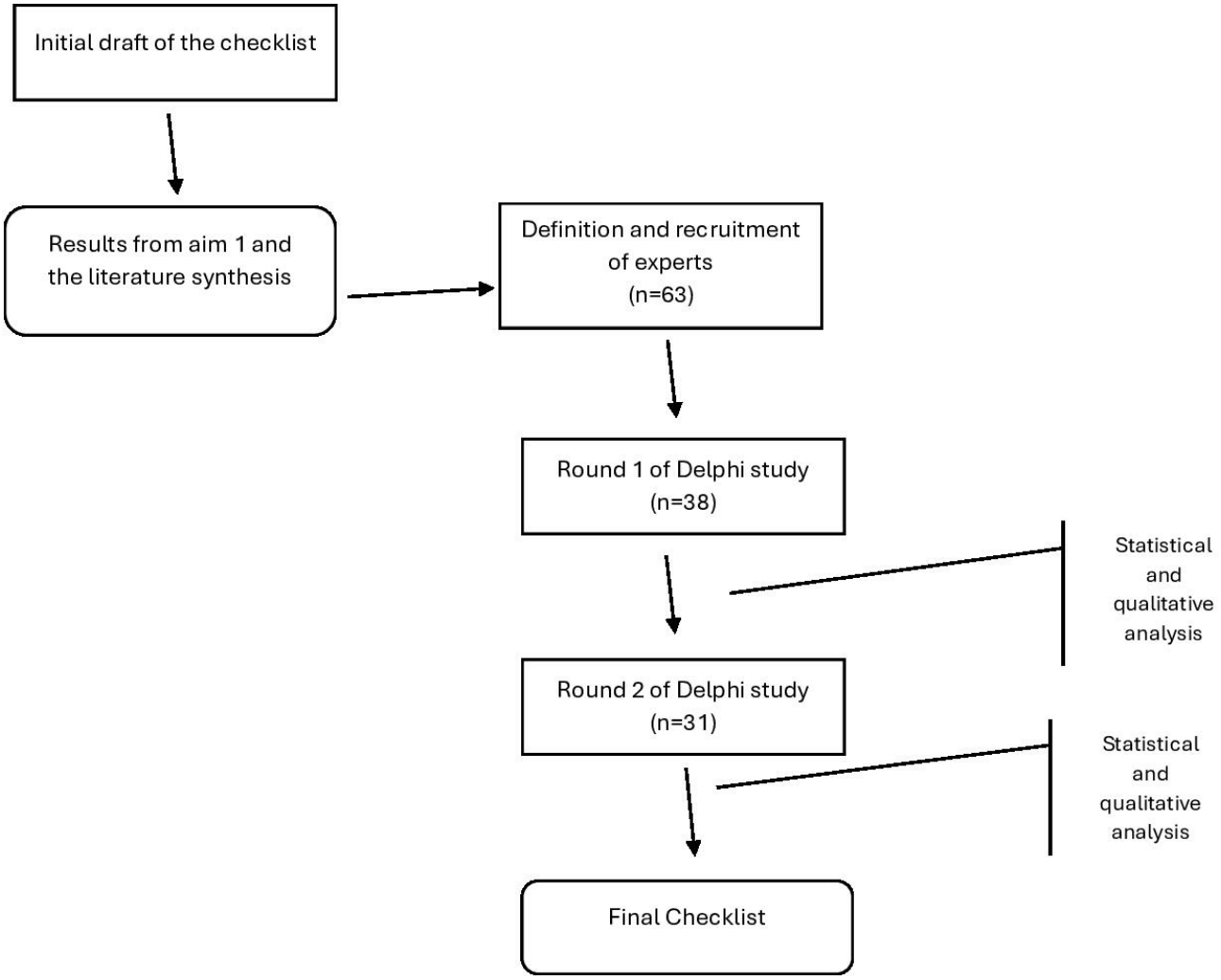
Participants who agreed to take part accessed the first round of the Delphi survey through a link in the email, which led to the consent form and survey. Data collection was done using Google Forms. To avoid bias, the panelists remained anonymous to each other throughout the process.

The preliminary survey comprised 45 questions designed to assess the relevance of each checklist item to the AI system's design and development process. A Likert scale from 1 ("Not Relevant") to 5 ("Highly Relevant") was used, along with open-ended comment fields for feedback and suggestions. The checklist was organized into four stages of AI system development: (1) planning, (2) design, (3) development, and (4) proposed implementation. Each stage aligned with 1 of the 6 domains in our framework.

After completion of the preliminary survey, the results were analyzed to assess the level of consensus among panelists. Based on the analysis, along with participants' feedback and comments, the checklist was revised and updated for the second round of the Delphi process. All the initial panelists were also invited for the second round even if they missed the first. This approach was based on the study by Boel et al [13], which showed that inviting panel members who missed a previous round to a subsequent round led to better representations of opinions and reduced the chances of false consensus while not influencing the outcome. The results of the analysis and feedback were added to the questionnaire for the second round. The whole process is highlighted in Figure 1.

Questions rated 4 or higher were classified as "relevant" to streamline the analysis. At the same time, those rated 3 or lower were deemed "irrelevant." This categorization facilitated a more efficient evaluation of the panelists' responses. Descriptive statistics were used to analyze the results of each round, along with an analysis of the IQR for each question. In determining the threshold for consensus among panelists, a mean score of 0.8 (representing 80% agreement) was established a priori as the benchmark. Questions with a mean score above 0.8 and an IQR of 0 were deemed to have consensus among the participants. Lastly, the Cronbach α reliability coefficient was calculated to evaluate the interitem reliability. The qualitative data collected during each round were analyzed using inductive content analysis. Quantitative analyses were conducted using the Python programming language in JupyterLab for Windows (Project Jupyter).

Figure 1. The process of developing the checklist.



Ethical Considerations

This study was conducted in accordance with institutional ethical guidelines for research involving human subjects and was approved by the University of Illinois Chicago Institutional Review Board under protocol STUDY2023-0535-MOD003. Participants provided informed consent, ensuring they were aware of the study's purpose, procedures, potential risks, and their right to withdraw at any time. All data collected were either anonymized or deidentified to protect participant privacy, with strict safeguards in place to ensure confidentiality. Additionally, no financial or material compensation was provided to participants in this Delphi study, and participation was entirely voluntary.

Results

Literature Synthesis

The literature search identified 20 studies [1,3,7,14-30] that proposed a framework, guideline, or approach for the design, development, implementation, or evaluation of AI for clinical use cases (Figure 2). A total of 14 (65%) of these addressed specific areas in the AI development cycle, from design to maintenance and management, while some cut across every aspect of the cycle. The results of the literature search were synthesized with the primary research and connected using a mind map to arrive at the domains of the Clinical AI Sociotechnical Framework (CASoF), which is a sociotechnical framework to support the planning, design, development, and proposed implementation of AI systems to help better plan and predict the likely success of the AI system (Figure 3).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews for Meta-Analyses) flowchart.

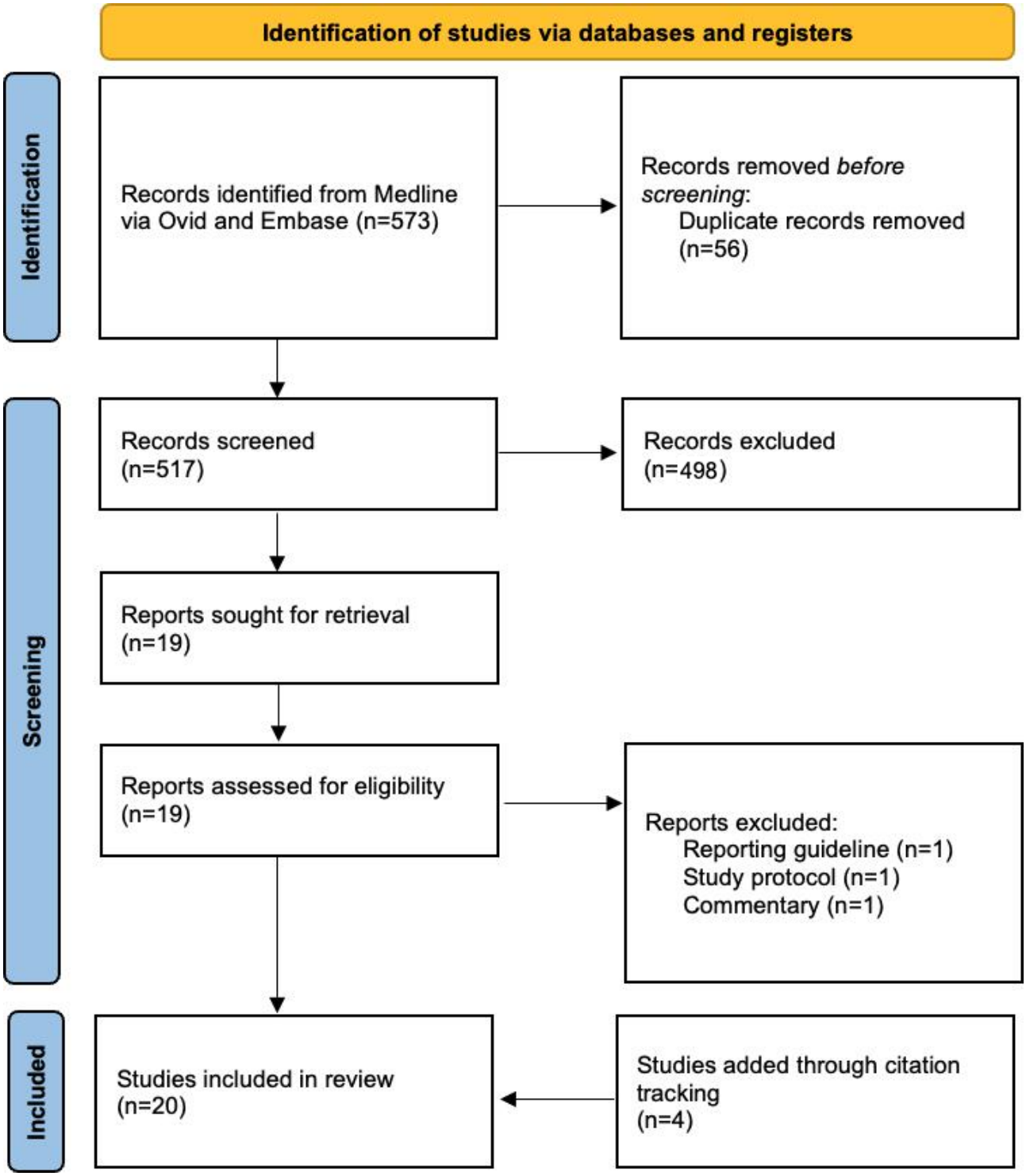
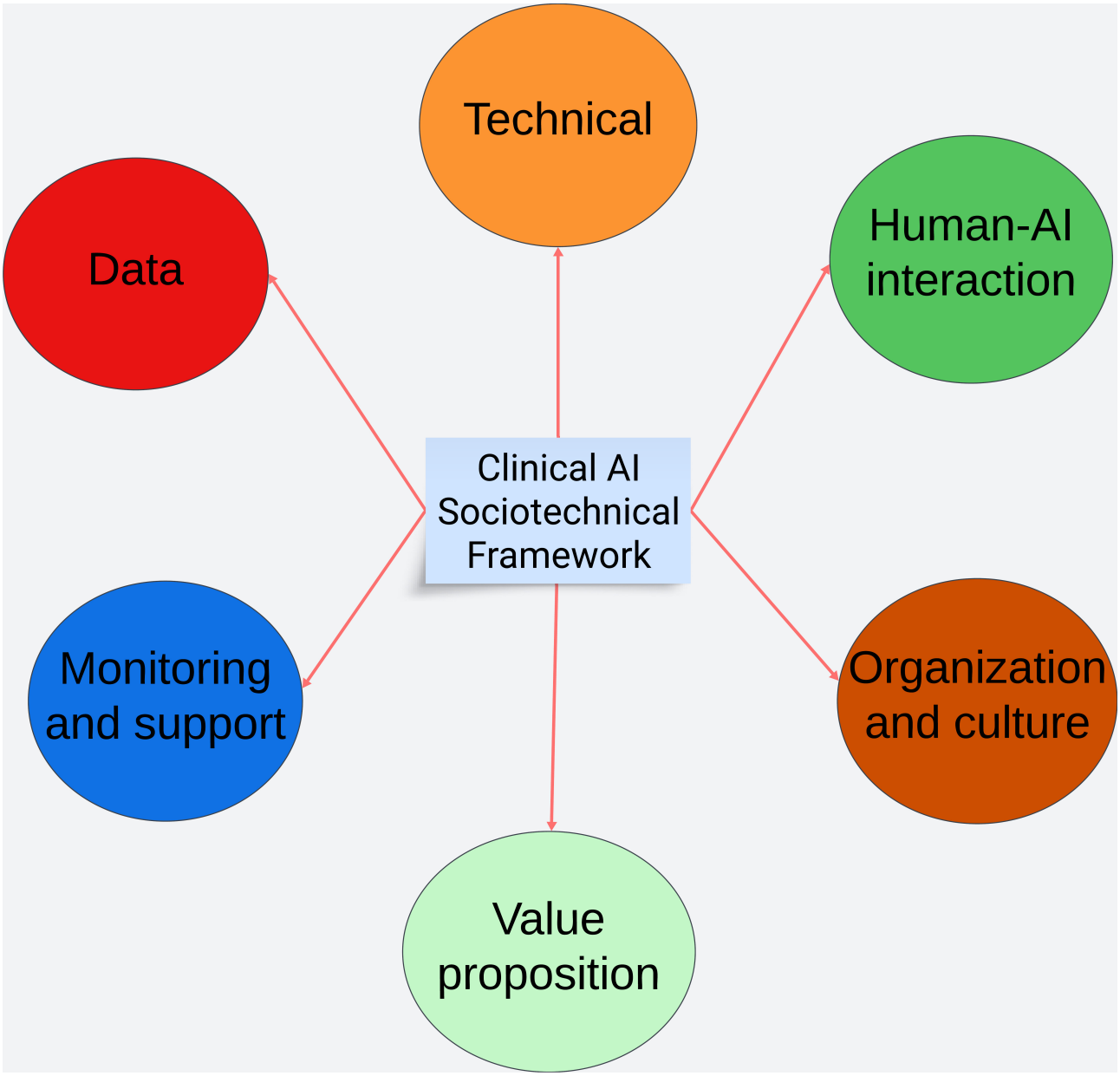


Figure 3. The Clinical AI Sociotechnical Framework. AI: artificial intelligence.



The Modified Delphi Study

Based on the CASoF, the first draft of the checklist was developed, which was shared with a team of panelists for evaluation and review using a Delphi approach. A total of 65 panelists were recruited: 21 (32%) doctors, 10 (15%) health care experts or researchers, 9 (12%) AI researchers, 4 (6%) health informaticians, 4 (6%) nurses, and 18 (28%) other professionals. Of the 65 panelists invited to participate in the study, 35 (54%) of them completed the first round of Delphi. The initial checklist had 4 overall categories that corresponded to the 4 stages in the development and deployment process, with 15 subcategories that corresponded to the domains of the CASoF that were important in each of the stages. The stages were "Planning," "Design," "Development," and "Proposed Implementation." As part of the questionnaire, panelists were asked 2 open-ended questions at the end of each of the subcategories: "Would you reframe any of the questions above?" and "Are there questions that you would add or remove from this segment?" During the first round of the Delphi, panelists suggested multiple edits and additions to the checklist. This suggested editing included the need to reframe some of the questions to make them more appropriate and clearer for a checklist. In one of the subcategories, one panelist responded as follows:

> *The last question says, "data processing." That comes across as ambiguous. What does that refer to? who will be the audience for this survey? will they understand what that means? Are we trying to abstract curation, cleaning etc into abstraction?*

At the end of the survey, panelists were asked why they might not use the checklist, and some of the responses included the following:

> *I think the checklist is long. The challenge when you have checklists this long is that people tend to gloss over them and are not intentional about answering the questions in a detailed way.*

> *Might be helpful to shorted and make more actionable. eg, policies and procedures document has been completed versus have you considered a place for policies.*
>
> *The checklist is somewhat burdensome on the AI vendor and health system. I would cut the questions in half.*

These open-ended questions were analyzed using a content analysis approach to bring out the recurrent themes and perspectives shared by the panelists in reforming and improving the questionnaire. Quantitative analyses were done, which showed a high level of agreement and relevance across most questions. Descriptive analysis was done: the mean score for the relevance of the questions on the survey exceeded 0.8 on all but one, indicating that at least 80% of respondents found the questions pertinent to their work and the topic at hand. Furthermore, the IQR was calculated to be 0 for all questions except 3, highlighting a level of consensus among respondents. The consensus and the structure of the checklist are shown in Multimedia Appendix 1.

Based on the results, comments, and feedback from the panelists, the checklist was revised. The "Design" and "Development" stages were merged into a single stage, and the "People" and "Organization and Culture" domains were merged into a single domain. The "User Experience and Workflow" and "Clinical Utility" domains were merged to create a new domain called "Human-AI Interaction." The total number of questions was reduced from 45 questions to 34 questions to make it less cumbersome and more focused. These 34 questions were sent to all the registered panelists for a second round of the Delphi process. All the recruited panelists were included in the second round and invited to review the updated checklist. Quantitative analyses were done, which showed a high level of agreement and relevance across most questions. Descriptive analysis was done: the mean score for the relevance of the questions was more than 0.8 on all questions, indicating that at least 80% of respondents found the questions pertinent to their work and the topic at hand. Furthermore, the IQR was calculated to be 0 for all questions, highlighting a level of consensus among respondents. Based on the outcome of the Delphi study, a final checklist was outlined, with 1 more question added to make 35 questions in total (Table 1).

Table . Final draft of the Clinical AI[a] Sociotechnical Framework (CASoF) checklist.

Stage and domain		Questions
Planning		
	Value proposition and utility	• Have you outlined the expected impacts on patient outcomes? • Have you outlined its expected impact on care provider efficiency and outcomes? • Has any economic analysis been conducted for the AI system?
	Data	• Have you engaged in the use of any ethical data checklist during your data collection and preparation? • Have you engaged domain experts in the data preparation, cleaning, and engineering process? • Have you delineated an approach to maintain data quality, integrity, and security?
	People, organization, and culture	• Have you identified key stakeholders and their needs? • Have you identified potential resistance or barriers within the organization? • Are there strategies in place to facilitate and ensure end-user engagement in the design and development phase? • Do you have a good understanding of the culture within the institution and changes that might be needed?
Design and development		
	Technical	• Are you planning for hardware/software (EHR[b]) systems and requirements? • Have you conducted a real-world evaluation of the model? • Are you creating support documentation for users and management, eg, model details, explainability details, data details, metrics, manuals, etc? • Have you validated clinical accuracy and reliability? • Have you secured any required regulatory approval? • Have you taken active steps to mitigate against biased results?
	Human-AI integration	• Have you conducted a simulation with end users in real work system scenarios? • Have you evaluated if the outputs are clear and understandable for the users? • Have you implemented any patient and user safety measures? • Have you accounted for and evaluated existing clinical workflows? • Are you aligning the solution with existing protocols? • Have you assessed the impact on the delivery of clinical tasks? • Have you involved and tested with users? • Has any resistance to the use of the AI system been identified and addressed? • Are you developing strategies to ensure that the alerts from the AI system are relevant, timely, and not overwhelming, to avoid alert fatigue?

Stage and domain		Questions
	Data	• Have you tested your method on various types of data to make sure it works well in different situations? • Have you planned for data drift and shift (changes in the data over time)?
Proposed implementation		
	People, organization, and culture	• Have you ensured that this intervention aligns with the existing governance and regulatory frameworks of the organization? • Have you prepared necessary training/resources for end users? • Have you considered steps to help address end users' questions and alleviate their concerns?
	Technical	• Are you planning for pilot/silent tests? • Are you providing user tools for continuous validation and evaluation of the system?
	Monitoring and support	• Have you created a plan to evaluate the success of the implementation? • Have you planned for continuous user feedback on the system? • Have you planned for regular audits, reviews, and updates? • Have you planned for continuous education and support for users?

[a]AI: artificial intelligence.

[b]EHR: electronic health record.

Discussion

Principal Findings

We introduce the CASoF checklist, which is a checklist that was developed from the results of primary studies, a literature synthesis, and a modified Delphi process that involved multiple experts and health care professionals. The CASoF, based on its sociotechnical perspective, encompasses different existing frameworks by providing a structured overview of the critical issues related to the integration, validation, and operationalization of AI in health care. The CASoF offers a high-level approach to solving the translation and adoption problems bedeviling AI systems designed for clinical settings. The CASoF can be used singly or in combination with some of the other existing frameworks in evaluating AI systems. The Diagnostic Quality Model by Lennerz et al [16] and the Clinical Explainable AI Guidelines by Jin et al [17] address diagnostic quality and explainability within medical imaging. They provide structured methodologies that could refine the CASoF by integrating rigorous quality assessments and enhancing transparency in AI tools. The strengths of these frameworks lie in their focused criteria, which could synergistically enrich the CASoF's scope, ensuring that AI's clinical implementation is both effective and sociotechnically sound.

At the end of the Delphi study and reviews, 35 final questions were agreed on based on the consensus from the panel members. Adjustments and rearrangements were made to the sequence of questions based on the comments made as part of the feedback during the Delphi study. This is the first checklist that addresses sociotechnical factors across the phases of the AI cycle with a general approach that is not limited to any specific condition or use case in clinical care. The checklist aims to help ensure that AI solutions for clinical use cases are better built for impact, adoption, and success.

The checklist focuses on sociotechnical factors most relevant to achieving these outcomes. Some of the comments by the respondents highlighted how the high-level design of the checklist was a reason they might not use it; however, the checklist is intentionally made high level to make it as brief and less cumbersome as possible. One of the reasons it is high level is to make it easy to apply quickly by designers, developers, AI engineers, informaticians clinicians, and health care organization managers for the needed assessments; therefore, this checklist should be considered as a form of minimum guideline in the development and implementation of AI systems meant for clinical settings.

The checklist is divided into 3 stages corresponding to the phases of the AI development cycle. The domains are drawn from the domains of the CASoF, which are "Value Proposition," "Data," "Human-AI Interaction," "Organization and Culture," "Technical," and "Monitoring and Support" [31]. These domains are allocated to each stage based on their relevance to that stage. Some domains recur in different stages, like "Data," "Human-AI Interaction," "Organization and Culture," and "Technical." Other domains like "Value Proposition" and "Monitoring and Support" only appear in a single phase. Questions are outlined

under each domain based on the stage they belong to. The number of questions varies per stage and domain.

The questions must be answered with a "Yes," "No," or "Partially Done." Each stage is meant to be done before and after each corresponding phase of the development cycle, so that the development team knows what to plan for and later review what has been accomplished. The "Planning" stage addresses the decision and preparation phase of the project, which is where the groundwork is laid for the subsequent design of the system. This phase involves a value proposition assessment to determine if it ensures alignment with patients' and end users' benefits. It serves to help answer a "go or no go" question across the ethical, economic, and sociotechnical dimensions of the AI tool, which is part of what the "Planning" phase in the CASoF checklist is designed to support. While the Biological-Psychological, Economic, and Social checklist by Khan and Seto [32] covers the planning aspect of AI development, it does not go beyond that phase, which is a limitation in its application.

The "Design and Development" phase covers the necessary steps and factors to be considered while building the AI system, unlike the R-AI-DIOLOGY checklist, which, apart from being focused explicitly on AI systems in radiology, only addresses the technical aspects of the design and development phases [33]. The last part of the checklist helps to plan for implementation, focusing on organization, culture, and needed monitoring. The Translational Evaluation of Healthcare AI framework checklist offers an alternative to the CASoF checklist for implementation; however, its lack of sociotechnical components, such as human-AI integration, culture and organization, and monitoring and support, which are essential for adoption and maximizing utility, is a drawback [3]. The checklist's design, development, and preimplementation aspects can also be used by payers, buyers, and decision makers to evaluate AI systems being sold or proposed to them to ensure they have been well designed and built.

Most of the existing checklists in this domain are targeted at reporting medical research carried out in AI or machine learning [34]. The CASoF checklist differs from these and other existing checklists like the Technology, Organization, and People framework–based checklist, which is focused on helping digital leaders manage adoption challenges [35]. It has no domain that addresses how the AI is designed or built, unlike the CASoF checklist. The same goes for the DECIDE-AI (Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence) checklist, which is focused on reporting studies that involve the evaluation of AI systems during their implementation phase in the clinical setting [36]. While the CASoF checklist does not explicitly have questions that address ethical issues, there are multiple questions across different phases that raise the need to address the ethics of the data, patient outcomes, and the impact of the outputs of the AI system.

Enhancing the real-world impact of AI tools involves navigating a nuanced blend of technical and social elements. This process demands a strategic framework that guides the planning and preparation efforts throughout the AI tool's life cycle, from its initial conceptualization to its sustained application. The CASoF checklist is designed to support designers, developers, AI engineers, informaticians, clinicians, health care organization managers, and others in planning, monitoring, and evaluating AI systems being developed or sold to them for clinical care.

Limitations

While the primary research, literature synthesis, and Delphi technique offer a robust approach to the development of the framework and checklist for the development and integration of AI in the clinical setting, the real-world application could be more difficult and not as straightforward as the research might suggest. Therefore, there might be a need for continuous refinement of the CASoF through iterative feedback and broader engagement with more stakeholders. Future research should aim to include an even wider array of perspectives, particularly from underrepresented regions and specialties, to enhance the framework's comprehensiveness and applicability. The framework further encounters limitations in capturing the full spectrum of technical challenges, needs, and their implications across diverse health care contexts globally. Considering these constraints, the application of the framework will benefit from synergistic application with other existing frameworks.

Conclusion

The CASoF checklist offers an approach to bridge the gap between the technical aspects of AI and how they can be best planned to fit and work in the clinical setting, with a view to improving the impact it makes on clinical work and patient outcomes. It offers a structured strategy to mitigate challenges and obstacles in the development and implementation process. The CASoF offers an advancement over previous frameworks and approaches by holistically encapsulating the sociotechnical dimensions necessary for AI to thrive within the clinical space.

Acknowledgments

This work was supported in part by the Agency for Healthcare Research and Quality grant K01HS029042 (MES). This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

This checklist is available in an electronic format [37].

Authors' Contributions

AO contributed to the conceptualization, data collection, formal analysis, investigation, and methodology of the study. Additionally, AO drafted the original manuscript and participated in the review and editing process. JO contributed to writing, reviewing, and editing of the manuscript. MES provided formal analysis, project administration, and supervision and contributed to the review and editing of the manuscript. AB contributed to the conceptualization and formal analysis of the study, managed the project, provided resources, and supervised the research. He also participated in the review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of subcategories by domain for the first round of the Delphi study.
[DOCX File, 15 KB - xmed_v6i1e65565_app1.docx]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File, 86 KB - xmed_v6i1e65565_app2.docx]

References

1. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cogn Eng Decis Mak 2022 Dec;16(4):194-206. [doi: 10.1177/15553434221097357] [Medline: 36704421]
2. Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of Medicine. JAMA 2020 Feb 11;323(6):509-510. [doi: 10.1001/jama.2019.21579] [Medline: 31845963]
3. Reddy S, Rogers W, Makinen VP, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ Health Care Inform 2021 Oct;28(1):e100444. [doi: 10.1136/bmjhci-2021-100444] [Medline: 34642177]
4. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
5. Tarumi S, Takeuchi W, Chalkidis G, et al. Leveraging artificial intelligence to improve chronic disease care: methods and application to pharmacotherapy decision support for type-2 diabetes mellitus. Methods Inf Med 2021 Jun;60(S 01):e32-e43. [doi: 10.1055/s-0041-1728757] [Medline: 33975376]
6. Ben-Israel D, Jacobs WB, Casha S, et al. The impact of machine learning on patient care: a systematic review. Artif Intell Med 2020 Mar;103:101785. [doi: 10.1016/j.artmed.2019.101785] [Medline: 32143792]
7. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022 Jan 27;24(1):e32215. [doi: 10.2196/32215] [Medline: 35084349]
8. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020 Sep;26(9):1364-1374. [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]
9. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 7;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]
10. Owoyemi A, Okpara E, Salwei M, Boyd A. End user experience of a widely used artificial intelligence based sepsis system. JAMIA Open 2024 Dec;7(4):ooae096. [doi: 10.1093/jamiaopen/ooae096] [Medline: 39386065]
11. Taylor E. We agree, don't we? The Delphi method for health environments research. HERD 2020 Jan;13(1):11-23. [doi: 10.1177/1937586719887709] [Medline: 31887097]
12. Taylor E, Joseph A, Quan X, Nanda U. Designing a tool to support patient safety: using research to inform a proactive approach to healthcare facility design. In: Rebelo F, Soares M, editors. Advances in Ergonomics In Design, Usability & Special Populations: Part III: AHFE International; 2022. [doi: 10.54941/ahfe1001343]
13. Boel A, Navarro-Compán V, Landewé R, van der Heijde D. Two different invitation approaches for consecutive rounds of a Delphi survey led to comparable final outcome. J Clin Epidemiol 2021 Jan;129:31-39. [doi: 10.1016/j.jclinepi.2020.09.034] [Medline: 32991995]
14. Parasa S, Repici A, Berzin T, Leggett C, Gross SA, Sharma P. Framework and metrics for the clinical use and implementation of artificial intelligence algorithms into endoscopy practice: recommendations from the American Society for Gastrointestinal Endoscopy Artificial Intelligence Task Force. Gastrointest Endosc 2023 May;97(5):815-824. [doi: 10.1016/j.gie.2022.10.016] [Medline: 36764886]
15. Pham N, Hill V, Rauschecker A, et al. Critical appraisal of artificial intelligence-enabled imaging tools using the levels of evidence system. AJNR Am J Neuroradiol 2023 May;44(5):E21-E28. [doi: 10.3174/ajnr.A7850] [Medline: 37080722]

16. Lennerz JK, Salgado R, Kim GE, et al. Diagnostic Quality Model (DQM): an integrated framework for the assessment of diagnostic quality when using AI/ML. Clin Chem Lab Med 2023 Jan 25;61(4):544-557. [doi: 10.1515/cclm-2022-1151] [Medline: 36696602]
17. Jin W, Li X, Fatehi M, Hamarneh G. Guidelines and evaluation of clinical explainable AI in medical image analysis. Med Image Anal 2023 Feb;84:102684. [doi: 10.1016/j.media.2022.102684] [Medline: 36516555]
18. Chomutare T, Tejedor M, Svenning TO, et al. Artificial intelligence implementation in healthcare: a theory-based scoping review of barriers and facilitators. Int J Environ Res Public Health 2022 Dec 6;19(23):16359. [doi: 10.3390/ijerph192316359] [Medline: 36498432]
19. Daye D, Wiggins WF, Lungren MP, et al. Implementation of clinical artificial intelligence in radiology: who decides and how? Radiology 2022 Dec;305(3):555-563. [doi: 10.1148/radiol.212151] [Medline: 35916673]
20. Tsopra R, Fernandez X, Luchinat C, et al. A framework for validating AI in precision medicine: considerations from the European ITFoC consortium. BMC Med Inform Decis Mak 2021 Oct 2;21(1):274. [doi: 10.1186/s12911-021-01634-3] [Medline: 34600518]
21. Jha AK, Myers KJ, Obuchowski NA, et al. Objective task-based evaluation of artificial intelligence-based medical imaging methods:: framework, strategies, and role of the physician. PET Clin 2021 Oct;16(4):493-511. [doi: 10.1016/j.cpet.2021.06.013] [Medline: 34537127]
22. Truong T, Gilbank P, Johnson-Cover K, Ieraci A. A framework for applied AI in healthcare. Stud Health Technol Inform 2019 Aug 21;264:1993-1994. [doi: 10.3233/SHTI190751] [Medline: 31438445]
23. de Hond AAH, Leeuwenberg AM, Hooft L, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022 Jan 10;5(1):2. [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
24. Sendak MP, Ratliff W, Sarro D, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020 Jul 15;8(7):e15182. [doi: 10.2196/15182] [Medline: 32673244]
25. Assadi A, Laussen PC, Goodwin AJ, et al. An integration engineering framework for machine learning in healthcare. Front Digit Health 2022 Aug 4;4:932411. [doi: 10.3389/fdgth.2022.932411] [Medline: 35990013]
26. Hantel A, Clancy DD, Kehl KL, Marron JM, Van Allen EM, Abel GA. A process framework for ethically deploying artificial intelligence in oncology. J Clin Oncol 2022 Dec 1;40(34):3907-3911. [doi: 10.1200/JCO.22.01113] [Medline: 35849792]
27. Nagaraj S, Harish V, McCoy LG, et al. From clinic to computer and back again: practical considerations when designing and implementing machine learning solutions for pediatrics. Curr Treat Options Pediatr 2020;6(4):336-349. [doi: 10.1007/s40746-020-00205-4] [Medline: 38624409]
28. Bedoya AD, Economou-Zavlanos NJ, Goldstein BA, et al. A framework for the oversight and local deployment of safe and high-quality prediction models. J Am Med Inform Assoc 2022 Aug 16;29(9):1631-1636. [doi: 10.1093/jamia/ocac078] [Medline: 35641123]
29. Bazoukis G, Hall J, Loscalzo J, Antman EM, Fuster V, Armoundas AA. The inclusion of augmented intelligence in medicine: a framework for successful implementation. Cell Rep Med 2022 Jan 18;3(1):100485. [doi: 10.1016/j.xcrm.2021.100485] [Medline: 35106506]
30. Choudhury A. Toward an ecologically valid conceptual framework for the use of artificial intelligence in clinical settings: need for systems thinking, accountability, decision-making, trust, and patient safety considerations in safeguarding the technology and clinicians. JMIR Hum Factors 2022 Jun 21;9(2):e35421. [doi: 10.2196/35421] [Medline: 35727615]
31. Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. AI Ethics 2023 Feb;3(1):223-240. [doi: 10.1007/s43681-022-00195-z]
32. Khan WU, Seto E. A "Do No Harm" novel safety checklist and research approach to determine whether to launch an artificial intelligence-based medical technology: introducing the Biological-Psychological, Economic, and Social (BPES) framework. J Med Internet Res 2023 Apr 5;25:e43386. [doi: 10.2196/43386] [Medline: 37018019]
33. Haller S, van Cauter S, Federau C, Hedderich DM, Edjlali M. The R-AI-DIOLOGY checklist: a practical checklist for evaluation of artificial intelligence tools in clinical neuroradiology. Neuroradiology 2022 May;64(5):851-864. [doi: 10.1007/s00234-021-02890-w] [Medline: 35098343]
34. Zrubka Z, Gulacsi L, Pentek M. Time to start using checklists for reporting artificial intelligence in health care and biomedical research: a rapid review of available tools. Presented at: 2022 IEEE 26th International Conference on Intelligent Engineering Systems (INES); Aug 12-15, 2022; Georgioupolis Chania, Greece p. 000015-000020. [doi: 10.1109/INES56734.2022.9922639]
35. Tursunbayeva A, Chalutz-Ben Gal H. Adoption of artificial intelligence: a TOP framework-based checklist for digital leaders. Bus Horiz 2024;67(4):357-368. [doi: 10.1016/j.bushor.2024.04.006]
36. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022 May 18;377:e070904. [doi: 10.1136/bmj-2022-070904] [Medline: 35584845]
37. Owoyemi A. Clinical AI sociotechnical framework (casof). Beadaut, Inc. URL: https://bit.ly/CASOF [accessed 2025-01-23]

Abbreviations

AI: artificial intelligence
CASoF: Clinical Artificial Intelligence Societechnical Framework
CONSORT: Consolidated Standards of Reporting Trials
DECIDE-AI: Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by CN Hang, E Meinert, T Leung; submitted 19.08.24; peer-reviewed by Anonymous, Anonymous, Anonymous, K Thompson, S Saripalli, S Zaki; revised version received 10.11.24; accepted 28.11.24; published 20.02.25.

Please cite as:
Owoyemi A, Osuchukwu J, Salwei ME, Boyd A
Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study
JMIRx Med 2025;6:e65565
URL: https://xmed.jmir.org/2025/1/e65565
doi:10.2196/65565

The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: A Qualitative Study

Ajit Kerketta*, MHA; Raghavendra A N*, PhD

CHRIST (Deemed to be University), Hosur Road, Bhavani Nagar, Bengaluru, India

*all authors contributed equally

Corresponding Author:
Ajit Kerketta, MHA
CHRIST (Deemed to be University), Hosur Road, Bhavani Nagar, Bengaluru, India

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2023.04.12.23288461v1

Companion article: https://med.jmirx.org/2025/1/e70808

Companion article: https://med.jmirx.org/2025/1/e70059

Abstract

Background: Rural health care delivery remains a global challenge and India is no exception, particularly in regions with Indigenous populations such as the state of Jharkhand. The Community Health Centres in Jharkhand, India, are staffed by Indigenous workers who play a crucial role in bridging the health care gap. However, their motivation and retention in these challenging areas are often influenced by a complex mix of sociocultural and environmental factors. One such significant but understudied influencing factor is alimentation, or nutrition, in rural settings. Previous studies have identified several motivators, including community ties, cultural alignment, job satisfaction, and financial incentives. However, the role of alimentation in their motivation and retention in rural areas has not been sufficiently explored.

Objective: This study aims to explore how the strong bond with locally produced food products impacts the retention of Indigenous community health workers (CHWs) in Jharkhand, India, and shed light on a crucial aspect of rural health care workforce sustainability.

Methods: This study adopted a phenomenological research design to explore the lived experiences and perspectives of Indigenous CHWs in Jharkhand. A purposive sampling method was used to select CHWs who had worked in rural areas for at least five years. Data were collected through semistructured interviews, focusing on the participants' experiences of rural alimentation and how it influences their motivation and retention for rural health care. The interviews were audio recorded, transcribed, and analyzed using thematic analysis to identify common themes and patterns in their experiences related to nutrition and retention.

Results: The study revealed that rural alimentation plays a significant role in both the motivation and retention of CHWs in Jharkhand. CHWs who experienced consistent access to local food reported higher job satisfaction, better physical well-being, and a stronger commitment to their roles. It has also been perceived that consuming nutrient-dense food products decreases the risk of chronic illness among rural populations. Additionally, community support systems related to alimentation were found to be crucial in maintaining motivation, with many CHWs emphasizing the importance of local food availability and cultural ties. The findings suggest that improving access to organic nutrition can positively influence the retention of CHWs in rural areas.

Conclusions: Indigenous communities have unique food habits and preferences deeply rooted in agriculture and arboriculture. Their traditional eating practices are integral to their rich cultural heritage, with significant social, symbolic, and spiritual importance. This study highlights the critical role of rural alimentation in motivating and retaining CHWs in rural Community Health Centres in Jharkhand. Therefore, addressing organic versus conventional food in rural health care policies plays a vital role in improving the retention rates of CHWs. By recognizing the interconnectedness of nutrition and workforce sustainability, health care systems can better support Indigenous CHWs and continue delivering health care services.

(JMIRx Med 2025;6:e48346) doi:10.2196/48346

KEYWORDS

rural alimentation; community health workers; motivation; retention; rural health; rural nutrition; workforce

Introduction

Rural Health and Alimentation

Community health workers (CHWs) play a vital role in providing primary health care services to rural populations in low- and middle-income countries [1,2]. However, retaining them in rural areas is challenging, largely due to low motivation. One potential factor influencing their motivation and retention is access to a diverse and nutritious diet or rural alimentation [3]. Although the term "alimentation" has existed in the English language since the late 16th century, it is rarely used. In Latin-based languages like French, "alimentation" conveys a holistic view of how humans produce, procure, prepare, share, consume, and digest their food, encompassing human, technological, sociocultural, and environmental aspects [4].

Significance of Alimentation in Rural Health Systems

The term "rural alimentation" in this study refers to the food that Indigenous people produce, acquire, prepare, share, consume, and digest; it is intimately linked to their sociocultural and environmental surroundings. Indigenous CHWs are also among those who are devoted to these tastes and preferences and will find it difficult to give up their native cuisine. Rural food products that are fresh, pure, unadulterated, nutrient dense, and low in pesticides appeal to CHWs [5]. This experience has lured CHWs to continue serving in rural health centers in Jharkhand, India. However, the lack of local and traditional food in metropolitan cities has negatively impacted their motivation, causing many health workers to select rural health care jobs [6]. Previous studies indicate that the desire for nutritious food in the urban setting significantly affects their motivation, job satisfaction, and retention rates. For example, a study in Ethiopia showed that providing nutritious food to CHWs increased job satisfaction and reduced attrition rates [7]. Similarly, a study conducted in Malawi demonstrated that CHWs accessing local food products were less likely to leave their jobs [8]. Despite the potential impact of rural alimentation on CHWs' motivation and retention, there is limited research on this topic in Jharkhand. Therefore, the study seeks to explore the question: "How does access to diverse and organic food in rural Jharkhand influence the motivation and retention of Indigenous community health workers?"

The study also aims to explore how Indigenous CHWs in Jharkhand perceive the impact of rural alimentation on their motivation and retention. Investigating the connection between rural alimentation and the motivation and retention of Indigenous CHWs will provide valuable insights into the factors that influence their engagement and commitment. We also intend to share the findings with policy makers and health care stakeholders, invoking the implementation of policies that support the well-being of CHWs, promote local food, and attract adequate CHWs to rural Jharkhand.

Study Background

Jharkhand is an eastern Indian state with a population of 39 million spread across 79,714 square kilometers (2019 census). Out of the total population, 24.05% reside in urban areas, while 75.95% live in rural areas [9]. Agriculture and agroforestry products are the primary sources of livelihood. However, the younger generation is increasingly migrating to metropolitan cities in search of better, more sustainable living opportunities. Most state regions are characterized by hills, rugged terrain, lakes, and rivers, presenting significant challenges. While some areas have plains and level topography nestled within natural surroundings, socioeconomic difficulties and a lack of infrastructure make it challenging for CHWs to stay in these locations. Consequently, Jharkhand faces a severe shortage of health workforce [10,11]. Approximately 80% of health care workers are stationed in metropolitan cities catering to the 24.05% of the population residing in urban areas, while 20% health care workers serve the 75.95% population living in rural areas [11-13]. Additionally, the population's strong beliefs in spirit worship and reliance on local quacks and tantric practices for their ill health further contribute to the short supply of CHWs. This study aims to provide evidence-based insights into the factors that promote the retention of CHWs in rural areas.

Methods

Study Design

We used a qualitative case research design to help understand the perspectives, emotions, and behaviors of Indigenous CHWs and uncover their in-depth experiences [14]. This approach focuses on understanding the subjective meaning that drove CHWs to work in rural Jharkhand [15,16]. This study selected participants with 5 years of service records in the respective Community Health Centers (CHCs).

Ethical Considerations

The study is not a clinical trial and, therefore, does not require registration to establish safety and efficacy standards. Nevertheless, ethical approval was obtained from the Institutional Review Board of CHRIST (Deemed to be University), Bangalore, India (CU: RCEC/00371/11/22). Written informed consent was obtained from each participant before conducting the interview. To further ensure the privacy of the participants, all names were changed to pseudonyms during the transcriptions of the text. However, the interviewers (AK and RNA) know the actual names of the interview participants. Each participant received a fixed remuneration of US $5.75 after completion of the interview as an acknowledgment of their time and contribution.

Setting and Sample

The corresponding author randomly selected and visited 3 CHCs in the eastern districts of Jharkhand to pilot the survey. This visit played a key role in shaping the development of the research objective: to explore the impact of rural alimentation (local and traditional food systems) on the motivation and

retention of Indigenous CHWs in rural India, as well as to establish a suitable research framework. CHWs from these randomly selected CHCs participated by completing a self-validated questionnaire with open-ended questions. In the main study, 30 CHWs were selected; of these, 10, 12, and 8 CHWs from the respective CHCs met the study criteria. They had served more than five years, expressed willingness to continue residing in rural areas, and were government employees. The study adopted a purposive sampling technique, which helped obtain rich, detailed, and relevant data that influenced the motivation and retention of CHWs in Jharkhand [17]. The male and female respondents were selected irrespective of their rural and urban backgrounds.

Process of Data Generation

A total of 14 participants (4 male individuals and 10 female individuals) ultimately consented to participate in the interviews. However, 16 individuals declined, with some initially agreeing but later withdrawing due to hesitation from the novelty of such an interview process and discomfort with having their comments audio recorded. The participants were aged 30-60 years and expressed their desire to participate in individual, face-to-face or telephone interviews within 10 months. A follow-up interview was done after 4 and 6 months. Within 4 months, 8 interviews were conducted at the CHCs and 6 interviews were conducted in the home district of the reviewer [18,19]. Multimedia Appendix 1 shows the interview guidelines and questionnaire.

The round-1 interview was precise and relevant to the objectives mentioned above (in the *Setting and Sample* section) and, hence, did not require a reinterview of any participants. Interviews were conducted both face-to-face and remotely in Hindi, a language in which the authors are fluent and experienced in conducting qualitative case research. While consent was sought to audio record the interviews, many participants expressed unwillingness; as a result, the researchers took detailed notes instead.

The data were collected through individual, semistructured qualitative case research, with in-depth interviews conducted according to the established protocol matrix [20]. Questions regarding all main areas were posed, albeit in varying order. The interviews in the 4-month follow-up ranged between 6 and 37 minutes (average of 10 min), and interviews in the 6-month follow-up ranged between 5 and 13 minutes (average of 7 min).

Research Team and Reflexibility

AK (research scholar in human resource management, male, aged 40 y) and RAN (PhD in human resource management, male, aged 48 y) solely conducted the interviews. After the interviews, the corresponding author listened to the audio recordings, with several breaks between every audio recording, and transcribed them.

Analysis

We employed the general data analysis methodologies indicated below in the context of thematic analysis and read the texts multiple times to familiarize and better understand them [21]. Descriptive codes were then applied to data segments [22] relevant to the research question: how do the local food habits influence motivation and continuation of work, and do these factors impact decisions to remain in rural areas? This question was aligned with the objective of the study [23]. The coded data were grouped into themes using QDA Miner Lite software (Provalis Research), demonstrating the relationships between them and identifying themes using inductive methods. The themes were assessed and modified depending on their relevance to the data and the research topic, and they were blended as appropriate. After the themes were developed, they were further defined and given titles that accurately expressed their meanings [24]. Then, the researcher drafted the report. The thematic analysis involves a recursive process of moving back and forth between the data and the emerging themes. It is an iterative and reflexive process, requiring the researcher to consider their biases and assumptions throughout the analysis.

- In-depth investigation: This method provided an in-depth understanding of the study's objectives and phenomena [25]. It enabled the researchers to collect data from multiple sources and examine them comprehensively.
- Contextual analysis: The qualitative case research design allowed the researchers to focus on the social, cultural, economic, and political factors influencing the phenomenon [26].
- Interpretive analysis: This approach involved identifying themes and interpreting them in the context of the research objectives [21,27].
- Flexible design: The qualitative case research design is adaptable, allowing the researchers to evolve the design as data are collected and analyzed [28]. To explore complex and context-specific issues in real-life settings, the interview provided comprehensive insights into the CHWs' experiences, opinions, and perspectives regarding rural alimentation and its impact on their motivation and retention.

Results

Study Participants

We contacted 4 CHCs; however, the medical officer at one center declined to grant permission, citing concerns that the study might inadvertently violate government protocols. A total of 64 CHWs were contacted across the remaining 3 CHCs, who were directly appointed by the government and were under the age of 60 years. Table 1 shows that the majority of health workers were female, accounting for 52% (13/25), 61% (11/18), and 57% (12/21) across the 3 CHCs. Among the 14 participants, 71% (n=10) were female and 29% (n=4) were male. This sex disparity could be a potential area for further research, exploring why fewer male CHWs tend to remain in rural locations.

Table . Participants characteristics from CHCs[a] A, B, and C. This table combines the demographic and workplace preferences of health care workers across the 3 centers (A, B, and C).

Characteristics	Center A (n=25), n (%)	Center B (n=18), n (%)	Center (n=21), n (%)
Sex			
Male	12 (48)	7 (39)	9 (43)
Female	13 (52)	11 (61)	12 (57)
Age group (years)			
≤30	3 (12)	5 (28)	4 (19)
≥30	22 (88)	13 (72)	17 (81)
Residence			
Rural origin	25 (100)	18 (100)	21 (100)
Urban origin	0 (0)	0 (0)	0 (0)
Preferred workplace			
Rural area	8 (32)	12 (67)	15 (71)
Male	2 (25)[b]	4 (33)[b]	4 (27)[b]
Female	6 (75)[b]	9 (67)[b]	11 (73)[b]
Urban area	17 (68)	6 (33)	6 (29)
Male	3 (18)[c]	1 (17)[c]	2 (33)[c]
Female	14 (82)[c]	5 (83)[c]	4 (67)[c]

[a]CHC: Community Health Centre.

[b]Percentages are based the number of workers who preferred a rural workplace as the denominator.

[c]Percentages are based the number of workers who preferred an urban workplace as the denominator.

Data were analyzed by constructing a thematic analysis, identifying patterns and themes as guided by the research questions and objectives [24,29]. Emerging themes were verified through member checking to ensure accuracy and validity. This study offers a comprehensive understanding and valid representations [30] of the perspectives and experiences of CHWs staying in rural Jharkhand. The focus is on a specific area within the CHCs, which is predominantly tribal dominated. The analysis identified themes that offered insights into the barriers and facilitators affecting CHWs' access to and consumption of diverse and nutritious food, as well as how their food habits intersect with their roles as health promoters and caregivers.

The study explored three major themes, presented as main themes and their corresponding minor themes, as illustrated below. These themes reflect the perspectives, experiences, and perceptions of the Indigenous CHWs regarding their reasons for remaining in rural Jharkhand.

1. The impact of rural alimentation on Indigenous CHWs' motivation
2. Retention trends among Indigenous CHWs
3. Correlations between nutritional support and job satisfaction

Impact of Rural Alimentation on Indigenous CHWs' Motivation

Health and Nutrition

Local food, often known as "field to plate," plays a vital role in connecting Indigenous CHWs to rural health centers. Free from preservatives, pesticides, additives, and flavorings, this food comes straight from the field, offering freshness and abundance, which enhances both its quality and appeal.

> *Whenever people call me to see patients or visit their house, they offer me fresh produce from their farm and sometimes even "desi" (country) chicken for free. Where can you get such nutritious and healthy food in cities?* [Nurse BY, 4-month interview]

Community Engagement

A unique characteristic of Indigenous communities is their emphasis on communitarian living, characterized by strong bonds of sharing and caring for one another [31,32]. Farming serves as both a livelihood and a means of fostering community engagement and identity. Their connection to the land, local markets, and cultural festivals centered around regional cuisine strengthens their sense of belonging and deepens social ties within the community.

> *I visit the villages whenever I have time. During these visits, many people gather to sit and discuss the health and well-being of the community, and we motivate the children. On holidays and Sundays, I often take*

the village youth to the rivers for fishing. [Doctor BA, 4-month interview]

Work-Life Balance

The concern among these CHWs is their inability to manage their domestic chores, as distance limits regular visits to the family and family affairs. The opportunity to serve in their home town facilitates work-life balance and positively impacts their physical and mental health, reducing stress, increasing job satisfaction, and enhancing productivity [33].

Cultural Connection

Food habits often represent a deep cultural bond and sense of belonging [34]. It makes them feel a strong connection to their heritage and traditions through the food they grew up with, making it more appealing to remain in their hometown.

> *We gather together and prepare meals for every celebration in common for all young and old.* [Accredited social health activist PK, 4-month interview]

Retention Trends Among Indigenous CHWs

Recognition

In rural areas, doctors often receive deep respect and appreciation from the rural community. This sense of being valued and recognized enforced emotional fulfillment, encouraging CHWs to continue serving in these regions.

> *I feel like a celebrity, as wherever I go—whether it's the market, the community, or my workplace—people honour and respect me immensely.* [Doctor DM, 4-month interview]

Career Intentions

The state government implemented various strategies to encourage medical students to serve in rural areas, including career growth incentives such as district quotas for entrance into Bachelor of Medicine, Bachelor of Surgery programs; specialized training programs (eg, barefoot doctor training) for rural service; a 3-month community medicine internship in rural settings; government-sponsored quotas for postgraduate, diploma, and degree course selections; as well as the introduction of the Diplomate of National Board program with training conducted in district hospitals [35]. As a result, professional development opportunities, a supportive work environment, community integration, and work-life balance were factors that encouraged CHWs to choose rural areas [36].

> *Once I complete the rural posting then there is an opportunity for further professional growth and other career intentions.* [Doctor SM, 4-month interview]

Promote Local Food and Lifestyle

Access to local food and a lifestyle that aligns with their cultural values and traditions contribute to higher retention rates [37]. The availability of fresh, familiar foods and a slower pace of life compared to urban centers created a more appealing working environment for Indigenous CHWs.

> *When I eat food outside of my region, I face digestion problems. It may be because I am not used to spices and tastemakers. Our tribal food is simple and organic resulting in better health outcomes. Therefore, I prefer to be in rural areas.* [Nurse PK, 6-month interview]

Role of Cultural Beliefs and Practices

The study of sociocultural and economic factors that affected food consumption patterns in Arab countries demonstrates that the cultural beliefs and practices related to food significantly shaped dietary habits and food choices among rural communities [38]. However, in this study, CHWs reported that the ancient practices have a great impact and were driven by a need for local cuisine [39].

Correlations Between Nutritional Support and Job Satisfaction

Better Health and Productivity

Access to nutritional support ensures that health care workers in rural areas stay physically fit and energized, which enhances their job performance [40]. Knowing that their health and well-being will be supported through nutritious, locally sourced food can make rural postings more attractive.

> *I have observed that rural people generally don't suffer from chronic diseases, but rather face issues like accidents, sunburn, sunstroke, or water-borne diseases. We are fortunate to have access to nutritious and healthy food.* [Nurse SH, 4-month interview]

Incentives of Fresh, Organic, and Local Food

Rural areas offer access to fresh, organic, and culturally significant local food. The availability of healthy, farm-to-table meals can serve as a strong motivator for health care workers, making rural postings more appealing due to the unique lifestyle benefits they offer.

> *They don't pay me that time for the treatment I provide when I visit or am called to see patients. They often can't afford to pay, but they give me fresh vegetables, pulses, or fruits that they harvest on the spot. Where else, in urban areas, can you find such genuine incentives and fresh produce?* [Nurse RJ, 6-month interview]

Low Cost of Living

In rural areas, access to fresh, local food can be more affordable than in urban settings. The prospect of spending less on quality food while still enjoying a nutritious diet can make rural postings more financially appealing.

> *I go to a market with 1000 INR [US $15] and buy groceries for the next two to three weeks. Everything is so cheap and fresh in the village markets. Do you think the same in the cities?* [Lab technician AG, 4-month interview]

Ethnicity

The findings demonstrated that ethnicity substantially impacted the food habits of a person owing to traditions, social norms,

migration, and acculturation, which is evident within and outside India [41]. When one travels outside of their home country or region, this becomes quite apparent.

Discussion

Principal Findings

The study findings underscore the positive impact that rural alimentation plays in enhancing the contentment of CHWs and highlight the complex interplay between the rural work environment and the factors that drive their motivation [42]. The results indicate that CHWs with access to nutritious food experienced higher motivation and retention rates [43]. The objective of the study also aligns with previous research showing that psychological factors related to adopting a healthy diet can significantly boost life satisfaction and job motivation. In this study, CHWs expressed satisfaction and a sense of contentment with the availability and quality of food in rural areas. This is similar to the study conducted in Tanzania, which showed that access to nutritious food made CHWs more likely to remain in their positions for extended periods [44].

Role of Nutrition in Enhancing Job Satisfaction

Previous studies have determined that a healthy diet helps protect against many chronic diseases, reducing the risk of developing such conditions [45,46]. The availability of locally sourced, nourishing food enhances rural health care workers' motivation and urges providers and administrators to promote a local and healthy diet, which is a relatively simple and cost-effective strategy to improve CHW motivation and retention [47]. The impact of organic food remains to be determined; it helps reduce food safety risks such as pesticide residue and excessive additives [48].

While there is a strong correlation between nutrition and job satisfaction, few studies, especially in health-related fields, have explored this link. The job satisfaction and food habits of CHWs are largely influenced by their socioeconomic conditions and social and cultural practices. For Indigenous CHWs, local food products play a crucial role in maintaining their health and job satisfaction, which significantly impact their retention [49]. A balanced diet contributes to sustained energy and reduces feelings of fatigue and burnout, allowing workers to perform effectively, which enhances their satisfaction with their jobs. A study on nutra-ergonomics explores the relationship between workers, their work environment, and job satisfaction in connection to their nutritional status. It highlights nutrition as a key component of a safe and productive workplace, influencing physical and mental health, and contributing to long-term retention in their current roles [50].

Cultural and Community Ties

Indigenous peoples typically share a deep ancestral connection to their lands and natural resources. They possess distinct cultures, languages, beliefs, and knowledge systems and maintain strong bonds with their land, properties, and territories. Their unique heritage and traditions are central to their identity and way of life. Culturally and politically, they will find themselves out of place from the rest of society [48].

Impact of Nutritional Support

Nutrition contributes to many indicators of well-being, including maternal health, birth weight, child development, and oral health, and is an important determinant of chronic disease, which reduces life expectancy [51]. Inadequate nutritional intake is a major factor contributing to the burden of disease, and when individuals develop chronic conditions as a result, it often leads to significant out-of-pocket expenses for treatment [52,53].

Government Policy

To attract and retain health workers in rural areas, both the state and central governments have implemented several monetary and nonmonetary benefits:

- Monetary incentives: (1) Hard area allowances and provision of residential facilities; (2) flexible salary schemes, such as the "You Quote, We Pay" strategy, ensuring competitive compensation; and (3) performance-based increments of up to 50% [35,54,55]
- Nonmonetary benefits: (1) Professional development opportunities for doctors, nurses, and allied health workers, including upskilling programs; (2) educational incentives, such as additional National Eligibility cum Entrance Test (Postgraduate) marks—10% for each year of service in remote or difficult areas, up to a maximum of 30%; (3) special honorariums to encourage rural practice among specialists; and (4) reservation of 50% of medical diploma seats for in-service state government doctors who have served in remote or challenging areas

These policies address financial and professional needs, making rural health care roles more attractive and sustainable [35,54]

Implications of the Study

The study revealed several significant implications for the retention and motivation of CHWs in rural settings. It underscored that CHWs with access to nutritious and diverse local food products demonstrated higher motivation and retention rates.

First, enhancing the nutrition of CHWs leads to improved health outcomes within the communities they serve. Given their pivotal role in delivering primary health care services in resource-limited rural areas, ensuring the health and motivation of CHWs directly correlated with the quality of care provided to their communities. Second, addressing the nutritional requirements of CHWs assisted in mitigating the challenge of high turnover rates prevalent in rural areas. CHWs often encounter numerous obstacles that contribute to burnout and turnover, such as long working hours, inadequate remuneration, and inadequate support. Third, the study underscored the significance of tackling social determinants of health, including access to nutritious food, to enhance health care outcomes in underserved communities. By addressing these determinants, health disparities can be reduced, thereby fostering overall community health improvement.

Limitations of the Study

The study was conducted in a specific geographic area and focused on a particular group of CHWs. The study's lack of

robust statistical representation may affect the reliability and generalizability of the results.

Conclusion

The research investigated the relationship between rural alimentation and the motivation of Indigenous CHWs in Jharkhand, India. The findings demonstrated that the retention rates of Indigenous health care workers are positively influenced by their local cuisines and nutrition. Moreover, CHWs with access to organic and locally sourced food exhibited superior retention rates compared to Indigenous CHWs deployed in urban areas. This study also indicated that individuals often exhibit loyalty to their culinary preferences and dietary habits, which drives them to opt for local assignments. Consequently, rural sustenance plays a pivotal role in CHW retention, thereby enhancing the health outcomes of rural residents. In essence, the study underscored the significance of addressing the local diet requirements of CHWs to bolster their motivation and retention rates, consequently elevating the standard of health care services in rural settings. The implications drawn from the study hold crucial insights for policy makers and health care practitioners operating in similar contexts, offering valuable strategies for enhancing the retention and motivation of CHWs in rural areas.

Acknowledgments

The authors extend their heartfelt gratitude to Paul Lelen Hoakip, research scholar, for his invaluable insights and thoughtful contributions, which has greatly enhanced the quality of this paper. Special thanks also to Rev. Dr Ayres Fernandez and Rev. John Thekekara, PhD, for their support and guidance throughout the development of this work. Their expertise and encouragement were instrumental in refining the manuscript and ensuring its academic rigor.

Multimedia Appendix 1
Interview guidelines and questionnaire.
[DOCX File, 19 KB - xmed_v6i1e48346_app1.docx]

References

1. Bitton A, Ratcliffe HL, Veillard JH, et al. Primary health care as a foundation for strengthening health systems in low- and middle-income countries. J Gen Intern Med 2017 May;32(5):566-571. [doi: 10.1007/s11606-016-3898-5] [Medline: 27943038]
2. van Iseghem T, Jacobs I, Vanden Bossche D, et al. The role of community health workers in primary healthcare in the WHO-EU region: a scoping review. Int J Equity Health 2023 Jul 20;22(1):134. [doi: 10.1186/s12939-023-01944-0] [Medline: 37474937]
3. di Renzo L, Gualtieri P, Pivari F, et al. Eating habits and lifestyle changes during COVID-19 lockdown: an Italian survey. J Transl Med 2020 Jun 8;18(1):229. [doi: 10.1186/s12967-020-02399-5] [Medline: 32513197]
4. Aguilera JM. The concept of alimentation and transdisciplinary research. J Sci Food Agric 2021 Mar 30;101(5):1727-1731. [doi: 10.1002/jsfa.10823] [Medline: 32949020]
5. Warra AA, Prasad MNV. Chapter 16 - African perspective of chemical usage in agriculture and horticulture—their impact on human health and environment. In: Prasad MNV, editor. Agrochemicals Detection, Treatment and Remediation: Pesticides and Chemical Fertilizers: Butterworth-Heinemann; 2020:401-436. [doi: 10.1016/B978-0-08-103017-2.00016-7]
6. de Marco M, Thorburn S, Kue J. In a country as affluent as America, people should be eating: experiences with and perceptions of food insecurity among rural and urban Oregonians. Qual Health Res 2009 Jul;19(7):1010-1024. [doi: 10.1177/1049732309338868] [Medline: 19556404]
7. Asfaw A, Simane B, Hassen A, Bantider A. Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: a case study in Woleka sub-basin. Weather Clim Extrem 2018 Mar;19:29-41. [doi: 10.1016/j.wace.2017.12.002]
8. Palmer MA, Menninger HL, Bernhardt E. River restoration, habitat heterogeneity and biodiversity: a failure of theory or practice? Freshw Biol 2010 Jan 15;55(s1):205-222. [doi: 10.1111/j.1365-2427.2009.02372.x]
9. National Health Mission. 13th common review mission 2019. National Health Systems Resource Centre. 2019 URL: https://nhsrcindia.org/sites/default/files/2021-04/13th_common_review_mission-Report_2019_Revise.pdf
10. Government of India. Census of India 2011. 2011. URL: https://censusindia.gov.in/nada/index.php/catalog/1366/download/4478/Pesreport.pdf [accessed 2025-01-15]
11. Rural health statistics 2018-19. Ministry of Health and Family Welfare. 2019. URL: https://mohfw.gov.in/?q=reports/rural-health-statistics-2018-19 [accessed 2025-01-15]
12. Kok MC, Kane SS, Tulloch O, et al. How does context influence performance of community health workers in low- and middle-income countries? evidence from the literature. Health Res Policy Sys 2015 Mar 7;13:13. [doi: 10.1186/s12961-015-0001-3] [Medline: 25890229]
13. Jharkhand urban population. Population Census Data (india). 2011. URL: https://www.census2011.co.in/census/state/jharkhand.

html#:~:text=As%20per%20details%20from%20Census,are%2016%2C930%2C315%20and%2016%2C057%2C819%20respectively [accessed 2025-01-17]
14. Fossey E, Harvey C, McDermott F, Davidson L. Understanding and evaluating qualitative research. Aust N Z J Psychiatry 2002 Dec;36(6):717-732. [doi: 10.1046/j.1440-1614.2002.01100.x] [Medline: 12406114]
15. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative and Mixed Methods Approaches: Sage Publications, Inc; 2009. URL: https://collegepublishing.sagepub.com/products/research-design-6-270550 [accessed 2025-01-17]
16. Sutton J, Austin Z. Qualitative research: data collection, analysis, and management. Can J Hosp Pharm 2015;68(3):226-231. [doi: 10.4212/cjhp.v68i3.1456] [Medline: 26157184]
17. de Leeuw JA, Woltjer H, Kool RB. Identification of factors influencing the adoption of health information technology by nurses who are digitally lagging: in-depth interview study. J Med Internet Res 2020 Aug 14;22(8):e15630. [doi: 10.2196/15630] [Medline: 32663142]
18. Bebbington P, Wilkins S, Sham P, et al. Life events before psychotic episodes: do clinical and social variables affect the relationship? Soc Psychiatry Psychiatr Epidemiol 1996 Jun;31(3-4):122-128. [doi: 10.1007/BF00785758] [Medline: 8766457]
19. Ghai S, Dutta M, Garg A. Perceived level of stress, stressors and coping behaviours in nursing students. Indian J Posit Psychol 2014;5(1):60-65 [FREE Full text]
20. Jacobs K. Discourse analysis. In: Baum S, editor. Methods in Urban Analysis: Springer; 2021:151-172. [doi: 10.1007/978-981-16-1677-8_9]
21. Castleberry A, Nolen A. Thematic analysis of qualitative research data: is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
22. Tirandaz Z, Akbarizadeh G, Kaabi H. PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization. Measurement (Lond) 2020 Mar;153:107432. [doi: 10.1016/j.measurement.2019.107432]
23. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772. [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
24. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45. [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
25. Baškarada S. Qualitative case study guidelines. The Qualitative Report 2014;19(40):1-18. [doi: 10.46743/2160-3715/2014.1008]
26. Mbindyo P, Gilson L, Blaauw D, English M. Contextual influences on health worker motivation in district hospitals in Kenya. Implementation Sci 2009 Jul 23;4:43. [doi: 10.1186/1748-5908-4-43] [Medline: 19627590]
27. Coventry PA, Fisher L, Kenning C, Bee P, Bower P. Capacity, responsibility, and motivation: a critical qualitative evaluation of patient and practitioner views about barriers to self-management in people with multimorbidity. BMC Health Serv Res 2014 Oct 31;14(1):536. [doi: 10.1186/s12913-014-0536-y] [Medline: 25367263]
28. Cavaye ALM. Case study research: a multi‐faceted research approach for IS. Information Systems Journal 1996;6(3):227-242. [doi: 10.1111/j.1365-2575.1996.tb00015.x]
29. Peterson BL. Thematic analysis/interpretive thematic analysis. In: The International Encyclopedia of Communication Research Methods: Wiley; 2017:1-9. [doi: 10.1002/9781118901731.iecrm0249]
30. Willan J. Doing early childhood research. In: Early Childhood Studies: A Multidisciplinary Approach: Bloomsbury Publishing; 2017:353-371 URL: https://www.bloomsbury.com/ca/early-childhood-studies-9781137274021/ [accessed 2025-01-23] [doi: 10.1057/978-1-137-27402-1_18]
31. Coates J. In: Gray M, editor. Indigenous Social Work around the World: Towards Culturally Relevant Education and Practice: Routledge; 2016:1-5. [doi: 10.4324/9781315588360]
32. Smith-Morris C. Indigenous Communalism: Belonging, Healthy Communities, and Decolonizing the Collective: Rutgers University Press; 2020:25. [doi: 10.2307/j.ctvscxrb6.5]
33. Akintola O, Chikoko G. Factors influencing motivation and job satisfaction among supervisors of community health workers in marginalized communities in South Africa. Hum Resour Health 2016 Sep 6;14(1):54. [doi: 10.1186/s12960-016-0151-6] [Medline: 27601052]
34. Weber EU. Climate change demands behavioral change: what are the challenges? Social Research: An International Quarterly 2015;82(3):561-580. [doi: 10.1353/sor.2015.0050]
35. Ghosh K. Why we don't get doctors for rural medical service in India? Natl Med J India 2018;31(1):44-46. [doi: 10.4103/0970-258X.243416] [Medline: 30348926]
36. Agyapong VIO, Osei A, Farren CK, McAuliffe E. Factors influencing the career choice and retention of community mental health workers in Ghana. Hum Resour Health 2015 Jul 9;13(1):56. [doi: 10.1186/s12960-015-0050-2] [Medline: 26156234]
37. Schrank Z, Running K. Individualist and collectivist consumer motivations in local organic food markets. J Cons Cult 2018 Feb;18(1):184-201. [doi: 10.1177/1469540516659127]
38. Musaiger AO. Socio-cultural and economic factors affecting food consumption patterns in the Arab countries. J R Soc Health 1993 Apr;113(2):68-74. [doi: 10.1177/146642409311300205] [Medline: 8478894]

39. Chakona G. Social circumstances and cultural beliefs influence maternal nutrition, breastfeeding and child feeding practices in South Africa. Nutr J 2020 May 20;19(1):47. [doi: 10.1186/s12937-020-00566-4] [Medline: 32434557]
40. Karaferis D, Aletras V, Raikou M, Niakas D. Factors influencing motivation and work engagement of healthcare professionals. Mater Sociomed 2022 Sep;34(3):216-224. [doi: 10.5455/msm.2022.34.216-224] [Medline: 36310751]
41. Verbeke W, Poquiviqui López G. Ethnic food attitudes and behaviour among Belgians and Hispanics living in Belgium. British Food Journal 2005 Dec 1;107(11):823-840. [doi: 10.1108/00070700510629779]
42. Wronska MD, Coffey M, Robins A. Determinants of nutrition practice and food choice in UK construction workers. Health Promot Int 2022 Oct 1;37(5):daac129. [doi: 10.1093/heapro/daac129] [Medline: 36166265]
43. Ozano K, Simkhada P, Thann K, Khatri R. Improving local health through community health workers in Cambodia: challenges and solutions. Hum Resour Health 2018 Jan 6;16(1):2. [doi: 10.1186/s12960-017-0262-8] [Medline: 29304869]
44. Greenspan JA, McMahon SA, Chebet JJ, Mpunga M, Urassa DP, Winch PJ. Sources of community health worker motivation: a qualitative study in Morogoro Region, Tanzania. Hum Resour Health 2013 Oct 10;11:52. [doi: 10.1186/1478-4491-11-52] [Medline: 24112292]
45. Steinman L, Heang H, van Pelt M, et al. Facilitators and barriers to chronic disease self-management and mobile health interventions for people living with diabetes and hypertension in Cambodia: qualitative study. JMIR Mhealth Uhealth 2020 Apr 24;8(4):e13536. [doi: 10.2196/13536] [Medline: 32329737]
46. Cena H, Calder PC. Defining a healthy diet: evidence for the role of contemporary dietary patterns in health and disease. Nutrients 2020 Jan 27;12(2):334. [doi: 10.3390/nu12020334] [Medline: 32012681]
47. Mozaffarian D, Angell SY, Lang T, Rivera JA. Role of government policy in nutrition-barriers to and opportunities for healthier eating. BMJ 2018 Jun 13;361:k2426. [doi: 10.1136/bmj.k2426] [Medline: 29898890]
48. Chai D, Meng T, Zhang D. Influence of food safety concerns and satisfaction with government regulation on organic food consumption of Chinese urban residents. Foods 2022 Sep 22;11(19):2965. [doi: 10.3390/foods11192965] [Medline: 36230045]
49. Kendirkiran G, Batur B. The relationship between eating behavior and job satisfaction of academic staff. Int J Caring Sci 2022 Aug;15(2):825-836 [FREE Full text]
50. Shearer J, Graham TE, Skinner TL. Nutra-ergonomics: influence of nutrition on physical employment standards and the health of workers. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 2):S165-S174. [doi: 10.1139/apnm-2015-0531] [Medline: 27277565]
51. Alfred T, Corntassel J. Being Indigenous: resurgences against contemporary colonialism. Government and Opposition 2005;40(4):597-614. [doi: 10.1111/j.1477-7053.2005.00166.x]
52. Browne J, Hayes R, Gleeson D. Aboriginal health policy: is nutrition the 'gap' in 'Closing the Gap'? Aust N Z J Public Health 2014 Aug;38(4):362-369. [doi: 10.1111/1753-6405.12223] [Medline: 25091077]
53. Schembri L, Curran J, Collins L, et al. The effect of nutrition education on nutrition‐related health outcomes of Aboriginal and Torres Strait Islander people: a systematic review. Aust N Z J Public Health 2016 Apr;40(Suppl 1):S42-S47. [doi: 10.1111/1753-6405.12392] [Medline: 26123037]
54. Press Information Bureau, Ministry of Information and Broadcasting, Government of India. Revolutionizing healthcare: digital innovations in India's health sector. Press Information Bureau. 2024 Jan 15. URL: https://static.pib.gov.in/WriteReadData/specificdocs/documents/2024/jan/doc2024115298601.pdf [accessed 2025-01-15]
55. Silvestri DM, Blevins M, Afzal A, et al. Medical and nursing students' intentions to work abroad or in rural areas: an eight-country cross-sectional survey in Asia and Africa. Ann Glob Health 2015 Mar 12;81(1):52. [doi: 10.1016/j.aogh.2015.02.627]

Abbreviations

CHC: Community Health Centre
CHW: community health worker

Edited by A Schwartz; submitted 20.04.23; peer-reviewed by SK Thalari; revised version received 28.11.24; accepted 13.12.24; published 23.01.25.

Please cite as:
Kerketta A, A N R
The Impact of Rural Alimentation on the Motivation and Retention of Indigenous Community Health Workers in India: A Qualitative Study
JMIRx Med 2025;6:e48346
URL: https://xmed.jmir.org/2025/1/e48346
doi:10.2196/48346

Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review

Feryal Kurdi*, MD; Yahya Kurdi*, MD; Igor Vladimirovich Reshetov, MD, PhD

Department of Oncology, Radiotherapy and Plastic and Reconstructive Surgery, Sechenov University, Bolshaya Pirogovskaya, 6c1, Moscow, Russian Federation

*these authors contributed equally

Corresponding Author:
Feryal Kurdi, MD
Department of Oncology, Radiotherapy and Plastic and Reconstructive Surgery, Sechenov University, Bolshaya Pirogovskaya, 6c1, Moscow, Russian Federation

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.07.30.24311256v1

Companion article: https://med.jmirx.org/2025/1/e69705

Companion article: https://med.jmirx.org/2025/1/e68769

Abstract

Introduction: Breast cancer is the leading cause of morbidity and mortality worldwide. Accurate sentinel lymph node (SLN) mapping is crucial for staging and treatment planning in early-stage breast cancer. Indocyanine green (ICG) has emerged as a promising agent for fluorescence imaging in SLN mapping. However, comprehensive assessment of its clinical utility, including accuracy and adverse effects, remains limited. This scoping review aims to consolidate evidence on the use of ICG in breast cancer SLN mapping.

Objective: The objective of this scoping review is to evaluate the current literature on the use of ICG in SLN mapping for patients with breast cancer. This review aims to assess the accuracy, efficacy, and safety of ICG in this context and to identify gaps in the existing research. The outcomes will contribute to the development of further research as part of a PhD project.

Methods: Five electronic databases will be searched (PubMed, Embase, MEDLINE, Web of Science, and Scopus) using search strategies developed in consultation with an academic supervisor. The search strategy is set to human studies published in English within the last 11 years. All retrieved citations will be imported to Zotero and then uploaded to Covidence for the screening of titles, abstracts, and full text according to prespecified inclusion criteria. Patients with early-stage breast cancer (T1 and T2), selected T3 cases where the SLN biopsy is accurate, and those with clinically node-negative breast cancer will be included. The intervention criterion includes studies using ICG for SLN mapping and studies on the assessment of fluorescence imaging cameras. Citations meeting the inclusion criteria for full-text review will have their data extracted by 2 independent reviewers, with disagreements resolved by discussion. A data extraction tool will be developed to capture full details about the participants, concept, and context, and findings relevant to the scoping review will be summarized.

Results: The preliminary search began in December 2023. As of September 2024, papers have been screened and data are currently being extracted. Out of the 2130 references initially imported, 126 studies met the inclusion criteria after screening. The scoping review is expected to be published in January 2025.

Conclusions: Although ICG technology has been used for SLN mapping in patients with breast cancer, initial searches in 2022 revealed limited data on this technique's feasibility, safety, and effectiveness. At that time, preliminary search of Scopus, MEDLINE, Embase, and PubMed identified no current or forthcoming systematic reviews or scoping reviews on the topic. However, recent searches indicate a substantial increase in research and reviews, reflecting a growing interest and evidence in this area.

(JMIRx Med 2025;6:e66213) doi:10.2196/66213

KEYWORDS

indocyanine green; ICG; sentinel lymph node; breast cancer; fluorescence; axillary lymph node mapping; NIR; surgical planning; near-infrared

Introduction

Sentinel lymph node (SLN) biopsy plays a crucial role in staging and prognosis in breast cancer management. The SLN is the initial lymph node to which breast cancer cells are likely to metastasize, and the presence of cancer cells in the SLN indicates a higher likelihood of further metastasis to other lymph nodes and distant organs [1].

SLN biopsy involves injecting a tracer substance into the breast, which then migrates to the SLN. The SLN is then identified, excised, and examined for cancer cells. If the SLN is free of cancer cells, it suggests that the cancer has not spread to other lymph nodes, eliminating the need for additional lymph node dissection. Conversely, if the SLN contains metastases, further dissection is typically required [2].

Over the past 2 decades, SLN biopsy using blue dye and radiotracers has been established as the diagnostic standard of care for patients with early-stage breast cancer who have clinically negative lymph nodes [3,4].

However, these methods come with certain drawbacks, including the potential for allergic reactions to the blue dye and the necessity of nuclear medicine facilities for radiotracer injection and detection. In a cohort undergoing blue dye and radiotracer injection procedures, a small number of adverse reactions, such as skin tattooing and anaphylaxis, were reported [5].

In recent years, near-infrared (NIR) fluorescence imaging using indocyanine green (ICG) has emerged as an alternative approach for SLN mapping in patients with breast cancer. ICG, a fluorescent dye, is injected into the breast, which then migrates to the SLNs. A NIR camera detects the fluorescence emitted by ICG, enabling the surgeon to identify and excise the SLNs [6,7].

This technology offers several advantages over traditional methods, including enhanced visualization of SLNs, a lower risk of allergic reactions, and the elimination of the need for nuclear medicine facilities. Furthermore, ICG has an excellent safety profile [8-11].

The importance of this topic stems from the potential of ICG technology to enhance the accuracy and safety of SLN mapping in patients with breast cancer. Precise identification and removal of the SLN are crucial for accurate staging and prognosis. Inaccurate SLN identification can lead to unnecessary lymph node dissection, resulting in complications such as lymphedema and impaired arm function. Sampling a larger number of SLNs may increase the risk of upper limb lymphedema, sensory deficits, and reduced shoulder function.

Landmark trials have shown a significant difference in morbidity rates when comparing SLN biopsy to axillary dissection, with rates of 25% and 70%, respectively [3,12]. Recent studies have reported excising, on average, 2 nodes per patient, likely due to advancements in NIR technology and ICG fluorescence protocols [13-17]. Nevertheless, further research is essential to assess the long-term outcomes and cost-effectiveness of ICG technology compared to traditional methods.

Methods

Overview

The proposed scoping review will be guided by the JBI methodology for scoping reviews [18]. The search strategy aims to locate both published and unpublished articles. An initial limited search of PubMed, Embase, MEDLINE, Web of Science, and Scopus was undertaken to identify relevant articles on the use of ICG for SLN mapping in breast cancer. In consultation with an academic supervisor, the keywords in the titles and abstracts of relevant articles, as well as the index terms used to describe these articles, were used to develop a comprehensive search strategy for PubMed, Embase, MEDLINE, Web of Science, and Scopus (see Multimedia Appendix 1). This strategy, including all identified keywords and index terms, will be adapted for each included database. The articles sourced from all included sources of evidence will be exported into Zotero (Corporation for Digital Scholarship).

Only articles published in English will be included due to the language proficiency of the reviewers. Articles published since January 1, 2014, will be included to ensure relevance, aligning with the project's consideration of recent data and the ongoing advancements in SLN mapping techniques using ICG.

JBI Methodology for Scoping Reviews

The outcomes of the scoping review will inform and frame three subsequent pieces of work planned as part of a PhD project:

1. Prospective cohort study on the long-term outcomes of ICG in SLN mapping
2. Systematic review and meta-analysis of ICG for SLN mapping in breast cancer
3. Development of standardized clinical guidelines and protocols for the use of ICG in SLN mapping in patients with breast cancer

The Participants-Concept-Context framework for this scoping review defines (1) the participants as patients with early-stage breast cancer, (2) the concept as the use of ICG for SLN mapping in patients with breast cancer, and (3) the context as SLN mapping that is performed as part of breast cancer staging and treatment planning.

Review Questions

The review questions are as follows:

1. What do we know about the evaluation and integration of emergent evidence on the use of ICG for SLN mapping in patients with breast cancer into clinical practice and decision-making?
2. To what extent is emergent evidence on the feasibility, safety, and effectiveness of ICG for SLN mapping integrated into clinical guidelines and decision-making processes?
3. How is emergent evidence on the use of ICG for SLN mapping evaluated and incorporated into clinical guidelines and decision-making processes?

For the purposes of this scoping review, emergent evidence refers to new research findings on ICG for SLN mapping that

have emerged after market launch and have not yet been fully integrated into clinical guidelines and practice.

Eligibility Criteria

The eligibility criteria are as follows. Participants will include patients with early-stage breast cancer (T1 and T2) and selected T3 cases where SLN biopsy has been shown to be accurate. Concept will include the use of ICG for SLN mapping in patients with breast cancer, as well as the assessment of imaging techniques and devices used in conjunction with ICG for SLN mapping. Context will include clinical settings where SLN mapping is performed as part of breast cancer staging and treatment planning.

This scoping review will consider both experimental and quasi-experimental study designs, including controlled before-and-after studies and controlled interrupted time-series studies. In addition, analytical observational studies including prospective and retrospective cohort studies, case-control studies, and analytical cross-sectional studies will be considered for inclusion. This review will also consider descriptive observational study designs such as descriptive cross-sectional studies for inclusion. Qualitative studies that focus on qualitative data will be considered for inclusion.

Following the search, all identified articles will be exported into Zotero. Then, the remaining articles will be uploaded into Covidence (Veritas Health Innovations Ltd). Titles and abstracts will then be screened by the lead author against the inclusion criteria for the scoping review. Potentially relevant articles will be retrieved in full and included in Covidence. The full text of these articles will be assessed in detail against the inclusion criteria by 2 independent reviewers. Reasons for the exclusion of sources of evidence at the full-text stage that do not meet the inclusion criteria will be recorded and reported in the scoping review. Any disagreements that arise between the reviewers at each stage of the selection process will be resolved through discussion or with an additional reviewer. The results of the search and the study inclusion process will be reported in full in the final scoping review and presented in a PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist, as extracted from Covidence (see Multimedia Appendix 2) [19].

Data will be extracted from all articles included in the scoping review by 2 independent reviewers, using a data extraction tool developed by the lead reviewer and piloted with about 15 articles to refine and improve it. The data extracted will include specific details about the participants, concept, context, study methods, and key findings relevant to the scoping review questions and will be imported into either Covidence or Microsoft Excel.

A draft extraction form is provided (see Multimedia Appendix 3). The draft data extraction tool will be modified and revised as necessary during the process of extracting data from each included article. Modifications will be detailed in the scoping review. Any disagreements that arise between the reviewers will be resolved through discussion or with an additional reviewer. If appropriate, authors of articles will be contacted to request missing or additional data, where required.

The evidence presented will directly respond to the scoping review's objective and questions. The data will be presented graphically or in diagrammatic or tabular form. A narrative summary will accompany the tabulated and/or charted results and will describe how the results relate to the scoping review's objective and questions.

Results

The preliminary search began in December 2023. As of September 2024, papers have been screened and data are currently being extracted. Out of the 2130 references initially imported, 126 studies met the inclusion criteria after screening (see Multimedia Appendix 4). The scoping review is anticipated to be published in January 2025.

Discussion

The significance of SLN mapping using ICG technology in breast cancer lies in its potential to enhance accuracy and safety, reduce complications, and improve patient outcomes [20]. Although ICG technology has been used for SLN mapping in patients with breast cancer, initial searches in 2022 revealed limited data on the feasibility, safety, and effectiveness of this technique. At that time, a preliminary search of Scopus, MEDLINE, Embase, and PubMed identified no current or forthcoming systematic reviews or scoping reviews on the topic. However, recent searches indicate a substantial increase in research and reviews, reflecting a growing interest and evidence in this area. Further studies are necessary to assess the long-term efficacy and cost-effectiveness of this technique and to identify the patient populations most likely to benefit.

The objective of this scoping review is to assess the extent of the literature on SLN mapping using ICG technology around the evaluation and integration of emergent evidence for benefits and harms; explore its feasibility, safety, and effectiveness in a larger cohort of patients with breast cancer; and provide guidance for clinical decision-making.

This scoping review could also identify specific patient populations, such as those with higher BMIs, who may benefit most from ICG technology. Additionally, patients who have undergone neoadjuvant therapy could be particularly advantageous candidates.

Factors such as the type of NIR cameras used, the learning curve for surgeons to become proficient with ICG for SLN detection, the availability of ICG and radioisotopes, the presence of nuclear medicine facilities, regional variations in ICG usage, and cost comparisons with the gold standard are also critical considerations in the broader adoption of this technology.

Limitations of this study include a lack of quantitative synthesis (ie, meta-analysis) of the results, which may limit the ability to draw strong conclusions. This scoping review serves as a foundational step toward a more comprehensive systematic review and meta-analysis guiding the clinical decision-making and the integration of ICG into standardized guidelines for SLN mapping in patients with breast cancer.

Acknowledgments

This scoping review is to contribute in part to a Doctor of Philosophy degree.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 14 KB - xmed_v6i1e66213_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File, 112 KB - xmed_v6i1e66213_app2.docx]

Multimedia Appendix 3
Data extraction instrument.
[DOCX File, 15 KB - xmed_v6i1e66213_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extraction flowchart.
[DOCX File, 79 KB - xmed_v6i1e66213_app4.docx]

References

1. McMasters KM, Tuttle TM, Carlson DJ, et al. Sentinel lymph node biopsy for breast cancer: a suitable alternative to routine axillary dissection in multi-institutional practice when optimal technique is used. J Clin Oncol 2000 Jul;18(13):2560-2566. [doi: 10.1200/JCO.2000.18.13.2560] [Medline: 10893287]
2. Gradishar WJ, Moran MS, Abraham J, et al. NCCN Guidelines® Insights: Breast Cancer, version 4.2023. J Natl Compr Canc Netw 2023 Jun;21(6):594-608. [doi: 10.6004/jnccn.2023.0031] [Medline: 37308117]
3. Mansel RE, Fallowfield L, Kissin M, et al. Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC Trial. J Natl Cancer Inst 2006 May 3;98(9):599-609. [doi: 10.1093/jnci/djj158] [Medline: 16670385]
4. Donker M, van Tienhoven G, Straver ME, et al. Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981-22023 AMAROS): a randomised, multicentre, open-label, phase 3 non-inferiority trial. Lancet Oncol 2014 Nov;15(12):1303-1310. [doi: 10.1016/S1470-2045(14)70460-7] [Medline: 25439688]
5. Nguyen CL, Zhou M, Easwaralingam N, et al. Novel dual tracer indocyanine green and radioisotope versus gold standard sentinel lymph node biopsy in breast cancer: the GREENORBLUE Trial. Ann Surg Oncol 2023 Oct;30(11):6520-6527. [doi: 10.1245/s10434-023-13824-6] [Medline: 37402976]
6. Polom K, Murawa D, Rho YS, Nowaczyk P, Hünerbein M, Murawa P. Current trends and emerging future of indocyanine green usage in surgery and oncology: a literature review. Cancer 2011 Nov 1;117(21):4812-4822. [doi: 10.1002/cncr.26087] [Medline: 21484779]
7. Liberale G, Vankerckhove S, Bouazza F, et al. Systemic sentinel lymph node detection using fluorescence imaging after indocyanine green intravenous injection in colorectal cancer: protocol for a feasibility study. JMIR Res Protoc 2020 Aug 14;9(8):e17976. [doi: 10.2196/17976] [Medline: 32554370]
8. van der Vorst JR, Schaafsma BE, Hutteman M, et al. Near-infrared fluorescence-guided resection of colorectal liver metastases. Cancer 2013 Sep 15;119(18):3411-3418. [doi: 10.1002/cncr.28203] [Medline: 23794086]
9. Hope-Ross M, Yannuzzi LA, Gragoudas ES, et al. Adverse reactions due to indocyanine green. Ophthalmology 1994 Mar;101(3):529-533. [doi: 10.1016/s0161-6420(94)31303-0] [Medline: 8127574]
10. Griffiths M, Chae MP, Rozen WM. Indocyanine green-based fluorescent angiography in breast reconstruction. Gland Surg 2016 Apr;5(2):133-149. [doi: 10.3978/j.issn.2227-684X.2016.02.01] [Medline: 27047782]
11. Benya R, Quintana J, Brundage B. Adverse reactions to indocyanine green: a case report and a review of the literature. Cathet Cardiovasc Diagn 1989 Aug;17(4):231-233. [doi: 10.1002/ccd.1810170410] [Medline: 2670244]
12. Giuliano AE, Hunt KK, Ballman KV, et al. Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. JAMA 2011 Feb 9;305(6):569-575. [doi: 10.1001/jama.2011.90] [Medline: 21304082]

13. Fleissig A, Fallowfield LJ, Langridge CI, et al. Post-operative arm morbidity and quality of life. results of the ALMANAC randomised trial comparing sentinel node biopsy with standard axillary treatment in the management of patients with early breast cancer. Breast Cancer Res Treat 2006 Feb;95(3):279-293. [doi: 10.1007/s10549-005-9025-7] [Medline: 16163445]
14. Ahmed M, Purushotham AD, Douek M. Novel techniques for sentinel lymph node biopsy in breast cancer: a systematic review. Lancet Oncol 2014 Jul;15(8):e351-e362. [doi: 10.1016/S1470-2045(13)70590-4] [Medline: 24988938]
15. Schaafsma BE, Verbeek FPR, Rietbergen DDD, et al. Clinical trial of combined radio- and fluorescence-guided sentinel lymph node biopsy in breast cancer. Br J Surg 2013 Jul;100(8):1037-1044. [doi: 10.1002/bjs.9159] [Medline: 23696463]
16. Ballardini B, Santoro L, Sangalli C, et al. The indocyanine green method is equivalent to the 99mTc-labeled radiotracer method for identifying the sentinel node in breast cancer: a concordance and validation study. Eur J Surg Oncol 2013 Dec;39(12):1332-1336. [doi: 10.1016/j.ejso.2013.10.004] [Medline: 24184123]
17. Abe H, Yamazaki K, Tokuda A, Ogawa M, Kawasaki M, Kameyama M. A novel approach for sentinel lymph node identification using fluororescence imaging and computed tomography lymphography in early-stage breast cancer patients. J Clin Oncol 2014 May 20;32(15_suppl):e12025-e12025. [doi: 10.1200/jco.2014.32.15_suppl.e12025]
18. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
19. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
20. Sugie T, Ikeda T, Kawaguchi A, Shimizu A, Toi M. Sentinel lymph node biopsy using indocyanine green fluorescence in early-stage breast cancer: a meta-analysis. Int J Clin Oncol 2017 Feb;22(1):11-17. [doi: 10.1007/s10147-016-1064-z] [Medline: 27864624]

Abbreviations

ICG: indocyanine green
NIR: near-infrared
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SLN: sentinel lymph node

Edited by S Tungjitviboonkun; submitted 06.09.24; peer-reviewed by Anonymous; revised version received 20.10.24; accepted 21.10.24; published 06.01.25.

Please cite as:
Kurdi F, Kurdi Y, Reshetov IV
Applications of Indocyanine Green in Breast Cancer for Sentinel Lymph Node Mapping: Protocol for a Scoping Review
JMIRx Med 2025;6:e66213
URL: https://xmed.jmir.org/2025/1/e66213
doi:10.2196/66213

Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development

Oguzhan Serin[1], MD; Izzet Turkalp Akbasli[1], MD; Sena Bocutcu Cetin[1], MD; Busra Koseoglu[1], MD; Ahmet Fatih Deveci[2], MSc; Muhsin Zahid Ugur[2], PhD; Yasemin Ozsurekci[3], MD

[1]Department of Pediatrics, Hacettepe University Medical School, Gevher Nesibe Avenue, Altindag, Ankara, Turkey

[2]Department of Health Information Systems, University of Health Sciences, Istanbul, Turkey

[3]Department of Pediatric Infectious Diseases, Hacettepe University Medical School, Ankara, Turkey

Corresponding Author:
Izzet Turkalp Akbasli, MD
Department of Pediatrics, Hacettepe University Medical School, Gevher Nesibe Avenue, Altindag, Ankara, Turkey

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.22.24303209v1

Companion article: https://med.jmirx.org/2025/1/e71100

Companion article: https://med.jmirx.org/2025/1/e71369

Companion article: https://med.jmirx.org/2025/1/e71098

Abstract

Background: Pneumonia is a leading cause of mortality in children aged <5 years. While machine learning (ML) has been applied to pneumonia diagnostics, few studies have focused on predicting the need for escalation of care in pediatric cases. This study aims to develop an ML-based clinical decision support tool for predicting the need for escalation of care in community-acquired pneumonia cases.

Objective: The primary objective was to develop a robust predictive tool to help primary care physicians determine where and how a case should be managed.

Methods: Data from 437 children with community-acquired pneumonia, collected before the COVID-19 pandemic, were retrospectively analyzed. Pediatricians encoded key clinical features from unstructured medical records based on Integrated Management of Childhood Illness guidelines. After preprocessing with Synthetic Minority Oversampling Technique–Tomek to handle imbalanced data, feature selection was performed using Shapley additive explanations values. The model was optimized through hyperparameter tuning and ensembling. The primary outcome was the level of care severity, defined as the need for referral to a tertiary care unit for intensive care or respiratory support.

Results: A total of 437 cases were analyzed, and the optimized models predicted the need for transfer to a higher level of care with an accuracy of 77% to 88%, achieving an area under the receiver operator characteristic curve of 0.88 and an area under the precision-recall curve of 0.96. Shapley additive explanations value analysis identified hypoxia, respiratory distress, age, weight-for-age *z* score, and complaint duration as the most important clinical predictors independent of laboratory diagnostics.

Conclusions: This study demonstrates the feasibility of applying ML techniques to create a prognostic care decision tool for childhood pneumonia. It provides early identification of cases requiring escalation of care by combining foundational clinical skills with data science methods.

(JMIRx Med 2025;6:e57719) doi:10.2196/57719

KEYWORDS

childhood pneumonia; community-acquired pneumonia; machine learning; clinical decision support system; prognostic care decision

Introduction

Pneumonia is responsible for 14% of all mortality in children aged <5 years and is included in World Health Organization (WHO) reports as the cause of death in 740,180 children in 2019 alone [1,2]. The Global Action Plan for the Prevention and Control of Pneumonia and Diarrhea, which was released by the WHO and UNICEF, aimed to reduce the mortality rate from pneumonia and diarrhea in children aged <5 years [2,3]. They have set targets that include vaccination, water and air sanitation, exclusively breastfeeding in the first 6 months, and eliminating pediatric HIV cases, along with appropriate pneumonia and diarrhea care.

It has been demonstrated that timely and accurate diagnosis of pneumonia and appropriately initiated treatment reduce mortality by up to 28% [4]. Diagnosis can often be difficult, since the clinical presentation of pneumonia in children is variable [5]. For this reason, the WHO has published the Integrated Management of Childhood Illness (IMCI) guidelines, which guide physicians in diagnosing, treating, and identifying danger signs of pneumonia [6]. While some cases of pneumonia are treatable with appropriate interventions, even low-cost or low-tech options [1], pneumonia remains a leading cause of morbidity and mortality, particularly in resource-limited countries and regions [2]. Managing high-risk populations continues to present significant challenges, especially in intensive care settings where patients often require advanced respiratory support. In addition, it has been shown that families seeking health services in resource-limited settings causes delays in providing appropriate treatment, leading to disease progression [7]. These highlight the need to improve medical care decisions, particularly in regions with limited resources, to reduce pneumonia-related morbidity and mortality.

Early and accurate recognition of patients who may require escalation of care to tertiary facilities is essential, particularly for those who will require mechanical ventilation or advanced respiratory support [8]. Predicting which patients will deteriorate is challenging due to the heterogeneous presentation of pneumonia, and clinical features such as hypoxia, respiratory distress, nutritional status, and comorbidities are critical markers that necessitate closer monitoring or transfer [9,10]. Prolonged duration of illness and failure to respond to initial treatments are also important as they may indicate inadequate treatment, misdiagnosis, or incorrect identification of potential pathogens, which can lead to the escalation of care [7,11].

Data science can provide actionable evidence for effective clinical intervention in pediatric diseases in the future [12] and can reduce inequality in health care [13]. Also, using big data and machine learning (ML) technologies is promising for childhood pneumonia in low- and middle-income countries (LMICs), especially patient-risk stratification for developing severe disease and mortality [14]. Because of their flexibility and high accuracy, ML models are used in medicine in the fields of prediction (prognostics) and classification (diagnostics) [12]. Additionally, the use of ML offers great promise for decision support in managing community-acquired pneumonia (CAP) in children, as demonstrated in recent studies. These include predicting intensive care unit needs [15], low-cost and noninvasive diagnostics for childhood pneumonia in resource-limited settings [16], supporting pathogen identification at admission only using basic clinical and laboratory features [11], and using natural language processing with ML for supporting clinical decisions on radiology reports [17].

It has been seen that the vast majority of data science studies on pneumonia aims to provide diagnostic support to the physician by processing radiological images [18]. However, diagnostic utilities are mostly unavailable in LMICs and primary care units. Therefore, physicians need prognostic support algorithms that distinguish between serious and nonserious cases without using advanced diagnostic equipment.

We aimed to develop an ML-based clinical decision support tool for childhood pneumonia that can be used by non–intensive care physicians, particularly those working in LMICs, in predicting the escalation of care and thereby ensuring the effective diagnosis and treatment of pneumonia, which is one of the 2025 goals of the WHO [1,3].

Methods

Case Definition and Patient Selection

Our study included pediatric patients who received inpatient treatment at Hacettepe University Medical School, a large, urban, tertiary, academic medical center in Ankara, Türkiye, between January 2014 and April 2020. The center serves a diverse range of pediatric patients from both urban and rural areas across the country, including those requiring advanced multidisciplinary care as well as those with less severe conditions. All patients were diagnosed with CAP based on the most recent IMCI guidelines, which provide a structured clinical framework focused on clinical features rather than advanced imaging or laboratory results [6,19]. Patients younger than 28 days of age (neonatal age), those older than 18 years, and those who had been hospitalized within the last 14 days were excluded.

The medical records of 437 patients were retrospectively examined by pediatricians, who encoded the candidate features from unstructured admission notes based on the IMCI guidelines (Tables 1 and 2). These variables were chosen based on their clinical value in clinical decision-making and their availability in primary care.

Table . Candidate features: clinical variables.

Clinical variables	Description
Age	Age in months at the time of admission
Weight (*z* score)	Standardized score based on Turkish children reference values [20], indirectly reflecting nutritional status
Gender	Biological sex (male or female)
Complaint period	Duration (days) from symptom onset to admission
Comorbidity	Presence of any significant underlying medical conditions, including congenital disorders, genetic syndromes, neuromuscular diseases, and chronic respiratory or cardiac issues
Recent antibiotics usage	Prescribed oral antibiotic use within the 14 days before admission, suggesting an inadequately treated infection or failure to respond initial care
Fever	Presence of elevated body temperature at admission
Cough	A key respiratory symptom at admission
Loss of appetite	Sign of systemic illness, reflecting impact on the patient's well-being
Respiratory distress	Presence of shortness of breath, rapid breathing (tachypnea), nasal flaring, or chest wall retractions at initial examination
Abnormal lung sounds	Auscultatory findings (eg, crackles or wheezing), indicative of pulmonary pathology at initial examination
Hypoxia	SaO_2[a] measured by pulse oximetry; hypoxia is defined as SaO_2 below 92% at initial examination
Level of care severity	Primary outcome; whether the patient requires pneumonia care at a tertiary care unit, including PICU[b] admission or respiratory support (oxygenation or ventilation), at any point during the hospital stay

[a]SaO_2: peripheral blood oxygen saturation.

[b]PICU: pediatric intensive care unit.

Table . Candidate features: laboratory variables.

Laboratory variables	Unit
Hemoglobin	Grams per deciliter (g/dL)
Leukocytes	Cells per liter ($\times 10^6$/L)
Lymphocytes	Cells per liter ($\times 10^6$/L)
Neutrophils	Cells per liter ($\times 10^6$/L)
Platelets	Cells per liter ($\times 10^9$/L)
C-reactive protein	Milligrams per liter (mg/L)
Albumin	Grams per deciliter (g/dL)
Sodium	Milliequivalents per liter (mEq/L)
Aspartate aminotransferase	Units per liter (U/L)
Alanine aminotransferase	Units per liter (U/L)

The primary outcome was the "level of care severity," scaled as severe or nonsevere. This categorization was made by physician-encoders based on whether the patient required referral to a tertiary care unit, using medical notes during the hospital stay. Children classified as severe included those admitted to the pediatric intensive care unit or those who required oxygenation or ventilation support at any time during the hospital stay.

Ethical Considerations

This study's design and procedures were approved by the Hacettepe University Clinical Research Ethics Committee with protocol GO-20/1182. Since this study is a retrospective analysis using previously collected data, informed consent was not required as per the ethics committee's approval. All data used in this study were deidentified before analysis to ensure participant privacy and confidentiality. No compensation was

provided to participants, as this study did not involve direct human participant recruitment.

Study Population

This study included 437 hospitalized patients with CAP, categorized into nonsevere (n=133, 30.4%) and severe cases (n=304, 69.6%). Demographic and clinical candidate variables, along with laboratory indices, were collected. Group comparisons were made using the Mann-Whitney *U* test for continuous variables and the χ^2 test for categorical variables, with significance set at $P<.05$. A summary of these characteristics and statistical comparisons are provided in Table 3.

Table . Characteristics of the study population by level of care severity (N=437).

Candidate variables	Nonsevere (n=133, 30.4%)	Severe (n=304, 69.6%)	Test statistic (*df*)	*P* value
Age (months), median (IQR)	44 (13 to 98)	23 (7 to 64.5)	16,602[a]	.003
Weight (*z* scores), median (IQR)	−0.57 (−1.4 to 0.45)	−0.7 (−2.5 to 0.4)	17,784[a]	.045
Complaint period (days), median (IQR)	4 (2 to 7)	4 (2 to 7)	19,274[a]	.44
Gender, n (%)			0.05[a]	.83
Male	68 (30.9)	152 (69.1)		
Female	65 (30)	152 (70)		
Comorbidity, n (%)	85 (28.7)	211 (71.3)	1.28[b] (1)	.26
Recent antibiotic usage, n (%)	40 (26.3)	112 (73.7)	1.87[b] (1)	.17
Fever, n (%)	100 (32.3)	210 (67.7)	1.68[b] (1)	.20
Cough, n (%)	115 (31.3)	253 (68.8)	0.50[b] (1)	.48
Loss of appetite, n (%)	37 (32)	80 (68)	0.11[b] (1)	.74
Respiratory distress, n (%)	43 (17.1)	208 (82.9)	49.30[b] (1)	<.001
Abnormal lung sounds, n (%)	102 (26.9)	277 (73.1)	16.70[b] (1)	<.001
Hypoxia, n (%)	20 (7.7)	240 (92.3)	156.82[b] (1)	<.001
Hemoglobin (g/dL), median (IQR)	11.6 (10.4 to 12.9)	11.6 (10.6 to 12.6)	20,022[a]	.87
Leukocytes ($\times 10^6$/L), median (IQR)	9900 (6800 to 14,600)	10,950 (8050 to 15,850)	17,837[a]	.05
Lymphocytes ($\times 10^6$/L), median (IQR)	2300 (1400 to 3700)	2800 (1900 to 4400)	17,039[a]	.01
Neutrophils ($\times 10^6$/L), median (IQR)	5285 (2700 to 9200)	6500 (3650 to 10,900)	17,645[a]	.045
Platelets ($\times 10^9$/L), median (IQR)	310 (225 to 386)	317.5 (230.5 to 425)	19,399[a]	.50
C-reactive protein (mg/L), median (IQR)	2.06 (0.79 to 7.67)	2.06 (0.83 to 7.35)	19,842[a]	.76
Albumin (g/dL), median (IQR)	3.9 (3.73 to 4.2)	3.9 (3.4 to 4.2)	17,121[a]	.01
Sodium (mEq/L), median (IQR)	136 (135 to 138)	136 (134 to 138)	19,657[a]	.64
Aspartate aminotransferase (U/L), median (IQR)	35 (26 to 42)	35 (28 to 50)	18,382[a]	.13
Alanine aminotransferase (U/L), median (IQR)	17 (12 to 26)	18 (13 to 29)	18,457[a]	.15

[a]Mann-Whitney *U* test.

[b]Chi-square test.

Data Preprocessing

Data preprocessing, analysis, visualization, and model setup were conducted using Python (version 3.12; Python Software Foundation). We used Python libraries such as *Pandas*, *NumPy*, *Matplotlib*, *Seaborn*, and *Plotly* for exploratory data analysis. For model development, the *PyCaret* library was used, which includes an unsupervised anomaly detection module to identify and handle anomalous data points. *PyCaret* also offers various preprocessing modules to iteratively handle missing data using the light gradient boosting machine (LightGBM) algorithm. In this method, missing values were treated as dependent variables

and predicted based on other available features, minimizing bias. Individual feature weights were applied during this process. Specifically, of the 415 cases, the following features had missing values: C-reactive protein (n=34, 8.2%), albumin (n=10, 2.4%), sodium (n=8, 1.9%), aspartate aminotransferase (n=16, 3.9%), and alanine aminotransferase (n=16, 3.9%). For numerical data, min-max scaling was applied, while categorical data were processed using one-hot encoding. These preprocessing steps ensured the dataset was well prepared for model training and validation.

Handling the Imbalanced Dataset

The balance of the dataset was assessed using Shannon entropy, yielding a value of 0.7, which indicates an imbalanced dataset. To address this, we applied Synthetic Minority Oversampling Technique (SMOTE)–Tomek, a refined variation of the widely recognized SMOTE. This approach combines oversampling of the minority class with the removal of overlapping samples from the majority class through Tomek links. So, the ratio of samples becomes 1:1. The *Imblearn* library was used for implementing data oversampling.

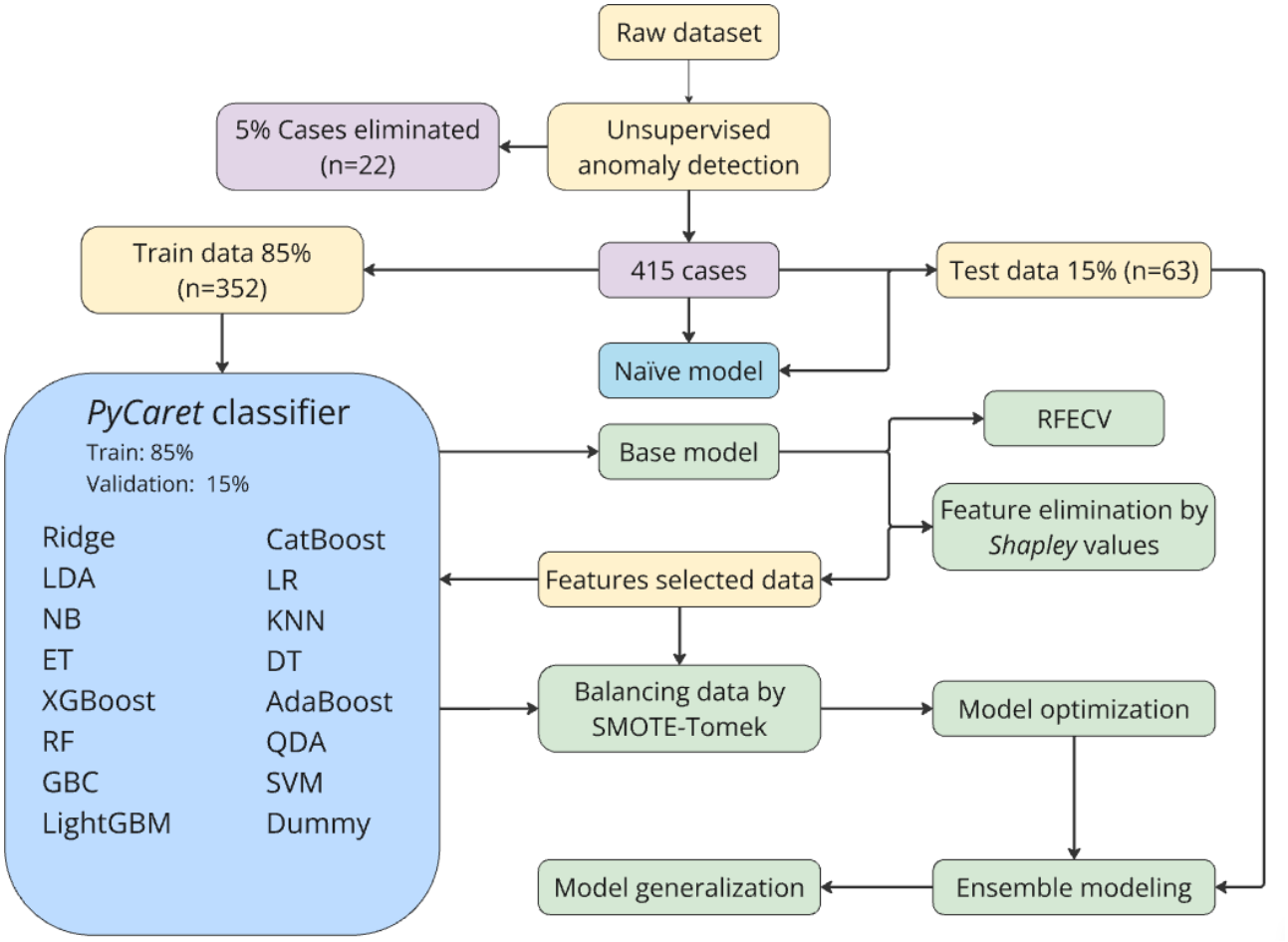
The dataset was split into two sets using the *train_test_split* method of the *SciKit-Learn* library. In the beginning, we allocated 5% of the general dataset as test data in order to prevent data leakage. The remaining 95% was split into training (352/415, 85%) and validation (63/415, 15%) sets.

Algorithms

PyCaret provides efficient implementations of state-of-the-art algorithms and is reusable among scientific disciplines. We used the *PyCaret* classifier module for classification, which includes the following models: ridge classifier, linear discriminant analysis, naïve Bayes, extra tree classifier, extreme gradient boosting (XGBoost), random forest, gradient boosting classifier, LightGBM, CatBoost classifier, logistic regression, k-neighbors classifier, decision tree, AdaBoost classifier, quadratic discriminant analysis, support vector machine with linear kernel, and dummy classifier.

In our work, we considered 10-fold cross-validation. While developing our model with *PyCaret* tools, we implemented the tuning function using the *Tune-Sklearn* library and the *hyper-band* optimization algorithm to obtain a set of best-performing parameters. For ensembling, we also used *PyCaret* classifier ensemble, stack, and blender methods. Ensembling methods have strong evidence that they can significantly enhance the accuracy of classifications [21]. After the optimization of parameters, in the last phase, we used the most common ensemble methods provided by the *PyCaret* library to further improve our model's performance (Figure 1).

Figure 1. The experimental setup: in this figure, we illustrate the experimental process of our models. Initially, we cleaned the data by identifying 5% of cases as abnormal data using unsupervised learning. We then split the data into a train set (85%) and a validation set (15%) using the *PyCaret* classifier model. The base model with the highest AUC-ROC value was the RF algorithm. Subsequently, we determined the optimal number of features as 18 using RFECV and selected the top 18 features based on Shapley values. We then balanced the dataset using the SMOTE-Tomek method and developed high-performing models. After optimizing the hyperparameters, we selected the best-performing model and created new models by using ensemble methods. In parallel, we developed a new model using only clinical findings for clinical prediction. AdaBoost: AdaBoost classifier; AUC-ROC: area under the receiver operator characteristic curve; CatBoost: CatBoost classifier; DT: decision tree; Dummy: dummy classifier; ET: extra tree classifier; GBC: gradient boosting classifier; KNN: k-neighbors classifier; LDA: linear discriminant analysis; LightGBM: light gradient boosting machine; LR: logistic regression; NB: naïve Bayes; QDA: quadratic discriminant analysis; RF: random forest; RFECV: recursive feature elimination with cross-validation; Ridge: ridge classifier; SMOTE: Synthetic Minority Oversampling Technique; SVM: support vector machine linear kernel classifier; XGBoost: extreme gradient boosting.



Feature Selection and Data-Reducing Methods

Feature selection is a process of one-by-one evaluation to determine which features are effective on the results within the dataset. Irrelevant or partially relevant features can negatively impact ML model performance and make the ML model learn based on irrelevant features. These methods are aimed at eliminating irrelevant features and keeping the strong features to reduce the dimension of the dataset. Recursive feature elimination is a feature selection method that fits a model and removes the irrelevant features until the specified number of features is reached. Recursive feature elimination with cross-validation (RFECV) aims to select the optimal number of features using permutation importance and recursive feature elimination. In this study, we used the *RFECV* module from *yellowbrick* library for selecting the optimum feature number. The Shapley additive explanations (SHAP) method is an innovative tool for explaining ML decision-making processes for datasets. The goal of the SHAP method is to present and explain the prediction with respect to the contribution of each feature to the predicted value. In RFECV, the features are ranked by a permutation importance measure. The SHAP algorithm was used for feature selection (Figure 2), as it provides more consistent and accurate importance values compared to the permutation approach. Ultimately, RFECV algorithms showed that 18 parameters are sufficient to explain nearly 90% of variances. Overall, 13 clinical and 5 laboratory variables were selected according to their SHAP values (Figure 2).

Figure 2. Feature selection: SHAP values are presented for the random forest classifier model with the highest AUC-ROC score in the dataset before feature selection, using the *SHAP* library's *plot_summary* module. The y-axis shows the importance of each feature, with the most important feature at the top and the least important at the bottom. The colors represent the contribution of each feature to the model's prediction. For example, features that have a large positive contribution to the prediction are shown in a warm color (eg, red), while features that have a large negative contribution are shown in a cool color (eg, blue). In this example, hypoxia is the most important attribute in the plot. The presence of hypoxia (hypoxia=1) causes the model to move closer to the target class, while its absence causes the model to move away from the target class. This predicts that hypoxia is an aggravating factor, while high levels of albumin have a protective effect for the target class. In summary, hypoxia is an adverse factor, and high albumin levels are protective. ALT: alanine aminotransferase; AST: aspartate aminotransferase; AUC-ROC: area under the receiver operator characteristic curve; CRP: C-reactive protein; SHAP: Shapley additive explanations.

Results

Study Population Characteristics

A comparison of the demographic and clinical characteristics between the nonsevere and severe groups is presented in Table 3. Of the 437 patients, 304 (69.6%) met the primary outcome, requiring the escalation of care. Patients in the severe care group were significantly younger, with a median age of 23 months compared to 44 months in the nonsevere level of care group (P=.003). Additionally, the severe group had lower weight z scores (P=.045).

Key clinical differences included higher rates of respiratory distress (208/304, 82.9% vs 43/133, 17.1%; P<.001), abnormal lung sounds (277/304, 73.1% vs 102/133, 26.9%; P<.001), and hypoxia (240/304, 92.3% vs 20/133, 7.7%; P<.001) in the severe group. In terms of laboratory findings, the severe group had higher leukocyte counts (P=.005), neutrophil counts (P=.045), and lymphocyte counts (P=.001). Albumin levels were slightly lower in the severe group (P=.01). No significant differences were observed between the groups in gender distribution (P=.83), comorbidities (P=.26), recent antibiotic use (P=.17), or C-reactive protein levels (P=.76).

Model Performances

In this section, we present a comparison of the performance of 16 different algorithms for raw and preprocessed datasets. We used various evaluation metrics such as accuracy, area under the receiver operator characteristic curve (AUC-ROC), recall, precision, F_1-score, Cohen κ, and Matthews correlation coefficient to assess model performance. To analyze model performance, all prediction experiments were conducted using 10-fold cross-validation. Subsequently, the models were optimized, and their performances were evaluated on a balanced dataset using SMOTE-Tomek and feature selection. The performances of the three models with the highest performance (CatBoost, XGBoost, and LightGBM) were evaluated by applying hyperparameter optimization and ensemble methods. Table 4 compares the results obtained with CatBoost, XGBoost, and LightGBM among the optimized and nonoptimized results, as well as the results of the combinations with the highest performance from the basic ensembling methods (ensembling, blending, and stacking methods). The highest AUC-ROC value was achieved by using optimized LightGBM as the meta-model in the stacking method.

Table . Comparative performance of machine learning models for the escalation of care prediction. Italicized values represent the highest scores for each column.

Model	Accuracy	AUC-ROC[a]	AUC-PRC[b]	Recall	Precision	F_1-score	Cohen κ	MCC[c]
CatBoost[d]	0.77	0.85	0.94	0.75	0.91	0.82	0.52	0.54
LightGBM[e,f]	0.80	0.87	0.96	0.79	0.92	0.85	0.58	0.59
XGBoost[f,g]	0.77	0.83	0.96	0.72	*0.94*	0.82	0.54	0.57
Ensembling[h]	0.77	0.86	0.95	0.72	*0.94*	0.82	0.54	0.57
Stacking[i]	0.80	*0.88*	0.96	0.79	0.92	0.85	0.58	0.59
Blending-1[j]	0.77	0.86	0.96	0.75	0.91	0.82	0.52	0.57
Blending-2[k]	*0.85*	0.84	0.96	*0.95*	0.85	*0.90*	*0.63*	*0.64*

[a]AUC-ROC: area under the receiver operating characteristic curve.

[b]AUC-PRC: area under the precision-recall curve.

[c]MCC: Matthews correlation coefficient.

[d]The performance of unoptimized CatBoost.

[e]LightGBM: light gradient boosting machine.

[f]The performance values obtained after optimization of XGBoost and LightGBM.

[g]XGBoost: extreme gradient boosting.

[h]The performance of the optimized LightGBM ensembling method, which achieved the highest results among CatBoost, XGBoost, and LightGBM algorithms.

[i]The performance of the model with optimized LightGBM as a meta-model in the stacking method, as it showed the highest performance.

[j]The combination of optimized LightGBM and XGBoost with higher performance in the blending method.

[k]Using the top-5, highest-ranked clinical features, the peak performance was realized by using a method that incorporated the optimized CatBoost, LightGBM, and XGBoost models.

In addition to the metrics reported in Table 4, we evaluated the performance of the *Blending-2* model using the precision-recall curve metric, which is particularly useful for imbalanced datasets. The precision-recall curve plot for this model, using the top-5 ranked clinical features, is provided in Multimedia Appendix 1. The model achieved a strong average precision-recall score of 0.96, further highlighting its robustness in handling imbalanced data.

Feature Importance

The optimized LightGBM in the model, developed with balanced and feature-selected data, was responsible for the

attainment of the highest performance. Upon evaluation of clinical features according to SHAP values, a ranking was established based on their feature importance scores, with the highest score being garnered by the top-5 clinical features (hypoxia, respiratory distress, age, *z* score of weight for age, and antibiotic usage before admission; Multimedia Appendix 2). The application of a workflow using these 5 features, as done previously, resulted in the highest accuracy performance (84%), which was achieved through the use of the ensemble method, incorporating the blending method of the optimized CatBoost, LightGBM, and XGBoost models.

Discussion

Pneumonia, the leading cause of childhood mortality, is also one of the most common causes of hospitalization [3,22]. It remains a significant global health burden, particularly in children aged <5 years, where timely and accurate clinical management is crucial for reducing mortality [8]. While prevention strategies are well documented, the clinical challenge lies in efficiently identifying patients who require escalated care. In this study, we present a contemporary approach to building an ML-based, prognostic care referral decision support tool that assists primary care physicians in determining where the case should be managed with an accuracy of more than 80%.

Today, there is widespread knowledge of the prevention, diagnosis, treatment, and management of complications in CAP, but due to resource limitations, it is not possible for all physicians and patients to benefit from this [14]. Recent advancements in medical informatics have the potential to reduce health care disparities and empower physicians in resource-limited settings [11-15], offering new hope for identifying high-risk populations and preventing mortality where current methods fall short.

The recent COVID-19 pandemic has impacted several medical fields, including the disruption of research practices by shifting researchers' focus and patient recruitment [23,24] and significantly reducing the incidence of non–COVID-19 pneumonia by preventing transmission [25-27]. In the current postpandemic state, non–COVID-19 childhood pneumonia remains a global health concern, especially in resource-limited settings according to the most recent reports [2], with respiratory infections likely to rise again as pandemic measures have already been eased [28]. Now, focusing back to reducing the mortality of CAP is critical to ensure pediatric pneumonia care benefits from recent advancements that COVID-19 provided [29,30]. This study, built primarily on prepandemic cases, provides a foundational context for future studies on CAP using ML in the postpandemic era.

Since March 2020, a substantial amount of data about COVID-19 have been published, including COVID-19–related artificial intelligence studies focused on pneumonia diagnosis by radiological findings [31]. However, pneumonia diagnosis is clinical, and routine chest radiographs are not necessary for the confirmation diagnosis [32] and do not improve outcomes [33]. In addition, chest radiography can be used only in inpatient settings to identify complications or evaluate response to treatment.

Although strong diagnostic support algorithms have been published in pneumonia-related studies in recent years, there is still a need for prognostic studies for pneumonia management [31]. Determining the severity of a disease or predicting its prognosis answers essential questions of physicians in medical decision-making, such as "Where should it be treated? Outpatient? ICU?" "Which therapy should I start? How long should I give it?" and "When should I discharge the patient? When should I call for control?" There are several studies and guidelines in the literature for severity assessment and prognosis prediction of pneumonia [9,10,34]. For the majority, mortality and the development of complications were the primary outcomes, and clinical, radiological, and laboratory variables are the key predictors. Yet, there is a limited number of studies predicting required referral to tertiary care based on basic clinical and laboratory features available in primary care settings [15].

This study reviewed important pneumonia prognostic predictors of children hospitalized in a major academic medical center. The primary outcome of interest was the level of care severity, classified as severe or nonsevere based on the need for pediatric intensive care unit admission or oxygen/ventilation support. The main objective of this study was not only to build the best model but also to answer the primary care physician's question: "Where should the case be managed?" Our model demonstrated promising predictive accuracy, with an AUC-ROC exceeding 0.85 and an accuracy of 77% to 88% (Table 4). The key clinical features identified—hypoxia, respiratory distress, age, *z* score of weight for age, and complaint period (Multimedia Appendix 2)—align with existing clinical guidelines, which emphasize the importance of respiratory and nutritional status in predicting disease severity [33-36].

In this study, we used SMOTE-Tomek, a method proven effective in medical tasks, to address class imbalance without losing valuable clinical information [37,38], which was essential given the significantly imbalanced and small sample–sized dataset. Additionally, we used RFECV and SHAP, both of which have been established as robust methods in previous studies [11,39,40], for feature selection. These techniques not only improved our model's performance but also allowed us to isolate the most clinically significant features (Figure 2, also see Multimedia Appendix 2), enabling clinicians to decide using their own skills without involving additional diagnostic tools.

The clinical application of a prognostic care decision model is particularly relevant in settings where early and accurate escalation of care is needed. For example, by focusing on these top-5 clinical features or using a decision support tool like ours, even less experienced primary care physicians could assess risk and anticipate tertiary care referrals without advanced diagnostics. Additionally, in emergency settings, these tools could assist in triaging patients to prioritize those needing immediate respiratory support or mechanical ventilation, allowing earlier interventions and more effective resource allocation—crucial for LMICs—potentially reducing morbidity and mortality.

One significant limitation of this study is its reliance on data from a single tertiary hospital (Hacettepe University), which

may limit generalizability. While the dataset includes patients referred from both urban and rural areas, the focus on a tertiary center introduces a selection bias, as most cases represent severe care levels (304/437, 69.6%). This is likely because less severe CAP cases are managed in primary or secondary care, not referred to tertiary centers, limiting the model's applicability in less severe cases. Additionally, the relatively small sample size of 437 patients limits the model's generalizability, as larger datasets are typically needed to optimize ML models and ensure robust performance across diverse populations. Expanding the dataset to include patients from multiple centers, especially primary and secondary care institutions, could improve the model's generalizability and applicability. Lastly, the retrospective nature of the data and the missing time frames of tertiary care unit transfers may not fully capture real-time clinical decision-making or the urgency of care decisions.

In conclusion, this study demonstrates the feasibility of developing an ML-based prognostic decision support tool for childhood pneumonia referral, with an accuracy of 77% to 88%. Incorporating foundational clinical skills for key prognostic predictors with advanced data science methods holds promise for improving pneumonia outcomes by accurately predicting the need for the escalation of care.

Acknowledgments

During the preparation of this work, the authors used OpenAI GPT-4o [41] to restructure sentences for enhanced readability, as they are not native English speakers. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Authors' Contributions

OS contributed to the creation of the work plan, interpretation of statistical analysis and machine learning algorithms, coinvestigation of the literature, and writing the revised manuscript. ITA contributed to the building of machine learning algorithms, coinvestigation of the literature, and writing the results and methods. SBC contributed to scanning patients from the hospital electronic health record system and encoding the attributes of the patients' data in the case report form ("Human Encoder-1"). BK contributed to the scanning patients from the hospital electronic health record system and encoding the attributes of the patients' data in the case report form ("Human Encoder-2"). AFD contributed to the building of the machine learning algorithms and optimizing the dataset. MZU contributed to the coding of advanced statistical and machine learning algorithms, and the creation of the clinical decision support system interface. YO contributed to the creation of the work plan, interpretation of statistical analysis, and gathering the team of investigators.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Precision-recall curve (PRC) for the blending model with top 5 features.
[PNG File, 21 KB - xmed_v6i1e57719_app1.png]

Multimedia Appendix 2
Shapley additive explanations (SHAP) values forward selection method.
[PNG File, 386 KB - xmed_v6i1e57719_app2.png]

References

1. Pneumonia in children. World Health Organization. 2022 Nov 11. URL: https://www.who.int/news-room/fact-sheets/detail/pneumonia [accessed 2024-10-01]
2. United Nations Inter-Agency Group for Child Mortality Estimation. Levels and trends in child mortality, report 2023. UNICEF. 2024 Mar 12. URL: https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2024/ [accessed 2024-10-01]
3. Qazi S, Aboubaker S, MacLean R, et al. Ending preventable child deaths from pneumonia and diarrhoea by 2025. development of the integrated Global Action Plan for the Prevention and Control of Pneumonia and Diarrhoea. Arch Dis Child 2015 Feb;100 Suppl 1:S23-S28. [doi: 10.1136/archdischild-2013-305429] [Medline: 25613963]
4. Sazawal S, Black RE, Pneumonia Case Management Trials Group. Effect of pneumonia case management on mortality in neonates, infants, and preschool children: a meta-analysis of community-based trials. Lancet Infect Dis 2003 Sep;3(9):547-556. [doi: 10.1016/s1473-3099(03)00737-0] [Medline: 12954560]
5. Shah SN, Bachur RG, Simel DL, Neuman MI. Does this child have pneumonia?: the rational clinical examination systematic review. JAMA 2017 Aug 1;318(5):462-471. [doi: 10.1001/jama.2017.9039] [Medline: 28763554]
6. World Health Organization. Handbook: IMCI integrated management of childhood illness. World Health Organization. 2005. URL: https://iris.who.int/handle/10665/42939

7. Ferdous F, Ahmed S, Das SK, et al. Pneumonia mortality and healthcare utilization in young children in rural Bangladesh: a prospective verbal autopsy study. Trop Med Health 2018 May 25;46:17. [doi: 10.1186/s41182-018-0099-4] [Medline: 29875615]
8. Shaima SN, Alam T, Bin Shahid A, et al. Prevalence, predictive factors, and outcomes of respiratory failure in children with pneumonia admitted in a developing country. Front Pediatr 2022 May 4;10:841628. [doi: 10.3389/fped.2022.841628] [Medline: 35601439]
9. Sonego M, Pellegrin MC, Becker G, Lazzerini M. Risk factors for mortality from acute lower respiratory infections (ALRI) in children under five years of age in low and middle-income countries: a systematic review and meta-analysis of observational studies. PLoS One 2015 Jan 30;10(1):e0116380. [doi: 10.1371/journal.pone.0116380] [Medline: 25635911]
10. McAllister DA, Liu L, Shi T, et al. Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. Lancet Glob Health 2019 Jan;7(1):e47-e57. [doi: 10.1016/S2214-109X(18)30408-X] [Medline: 30497986]
11. Chang TH, Liu YC, Lin SR, et al. Clinical characteristics of hospitalized children with community-acquired pneumonia and respiratory infections: Using machine learning approaches to support pathogen prediction at admission. J Microbiol Immunol Infect 2023 Aug;56(4):772-781. [doi: 10.1016/j.jmii.2023.04.011] [Medline: 37246060]
12. Bennett TD, Callahan TJ, Feinstein JA, et al. Data science for child health. J Pediatr 2019 May;208:12-22. [doi: 10.1016/j.jpeds.2018.12.041] [Medline: 30686480]
13. Zhang X, Pérez-Stable EJ, Bourne PE, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis 2017 Apr 20;27(2):95-106. [doi: 10.18865/ed.27.2.95] [Medline: 28439179]
14. Sheikh M, Jehan F. Using big data for risk stratification of childhood pneumonia in low-income and middle-income countries (LMICs): challenges and opportunities. EBioMedicine 2021 Dec;74:103740. [doi: 10.1016/j.ebiom.2021.103740] [Medline: 34916165]
15. Liu YC, Cheng HY, Chang TH, et al. Evaluation of the need for intensive care in children with pneumonia: machine learning approach. JMIR Med Inform 2022 Jan 27;10(1):e28934. [doi: 10.2196/28934] [Medline: 35084358]
16. Kanwal K, Khalid SG, Asif M, Zafar F, Qurashi AG. Diagnosis of community-acquired pneumonia in children using photoplethysmography and machine learning-based classifier. Biomed Signal Process Control 2024 Jan;87:105367. [doi: 10.1016/j.bspc.2023.105367]
17. Smith JC, Spann A, McCoy AB, et al. Natural language processing and machine learning to enable clinical decision support for treatment of pediatric pneumonia. AMIA Annu Symp Proc 2020 Jan 25;2020:1130-1139. [Medline: 33936489]
18. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med 2018 Nov 6;15(11):e1002683. [doi: 10.1371/journal.pmed.1002683] [Medline: 30399157]
19. Gera T, Shah D, Garner P, Richardson M, Sachdev HS. Integrated management of childhood illness (IMCI) strategy for children under five. Cochrane Database Syst Rev 2016 Jun 22;2016(6):CD010123. [doi: 10.1002/14651858.CD010123.pub2] [Medline: 27378094]
20. Neyzi O, Bundak R, Gökçay G, et al. Reference values for weight, height, head circumference, and body mass index in Turkish children. J Clin Res Pediatr Endocrinol 2015 Dec;7(4):280-293. [doi: 10.4274/jcrpe.2183] [Medline: 26777039]
21. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: a review. Healthcare (Basel) 2023 Jun 20;11(12):1808. [doi: 10.3390/healthcare11121808] [Medline: 37372925]
22. Jain S, Williams DJ, Arnold SR, et al. Community-acquired pneumonia requiring hospitalization among U.S. children. N Engl J Med 2015 Feb 26;372(9):835-845. [doi: 10.1056/NEJMoa1405870] [Medline: 25714161]
23. Shao JH, Yu KH, Chen SH. COVID-19-related disruptions in implementation of a randomized control trial: an autoethnographic report. Appl Nurs Res 2023 Aug;72:151698. [doi: 10.1016/j.apnr.2023.151698] [Medline: 37423680]
24. Sohrabi C, Mathew G, Franchi T, et al. Impact of the coronavirus (COVID-19) pandemic on scientific research and implications for clinical academic training - a review. Int J Surg 2021 Feb;86:57-63. [doi: 10.1016/j.ijsu.2020.12.008] [Medline: 33444873]
25. Kuitunen I, Artama M, Mäkelä L, Backman K, Heiskanen-Kosma T, Renko M. Effect of social distancing due to the COVID-19 pandemic on the incidence of viral respiratory tract infections in children in Finland during early 2020. Pediatr Infect Dis J 2020 Dec;39(12):e423-e427. [doi: 10.1097/INF.0000000000002845] [Medline: 32773660]
26. Chen M, Zhou Y, Jin S, et al. Changing clinical characteristics of pediatric inpatients with pneumonia during COVID-19 pandamic: a retrospective study. Ital J Pediatr 2024 Apr 23;50(1):84. [doi: 10.1186/s13052-024-01651-8] [Medline: 38650007]
27. Huang C. Pediatric non-COVID-19 community-acquired pneumonia in COVID-19 pandemic. Int J Gen Med 2021 Oct 27;14:7165-7171. [doi: 10.2147/IJGM.S333751] [Medline: 34737611]
28. Lastrucci V, Bonaccorsi G, Forni S, et al. The indirect impact of COVID-19 large-scale containment measures on the incidence of community-acquired pneumonia in older people: a region-wide population-based study in Tuscany, Italy. Int J Infect Dis 2021 Aug;109:182-188. [doi: 10.1016/j.ijid.2021.06.058] [Medline: 34216731]
29. Latif S, Usman M, Manzoor S, et al. Leveraging data science to combat COVID-19: a comprehensive review. IEEE Trans Artif Intell 2020 Sep 2;1(1):85-103. [doi: 10.1109/TAI.2020.3020521] [Medline: 37982070]

30. Hu S, Wang X, Ma Y, Cheng H. Global research trends in pediatric COVID-19: a bibliometric analysis. Front Public Health 2022 Feb 16;10:798005. [doi: 10.3389/fpubh.2022.798005] [Medline: 35252087]
31. Chumbita M, Cillóniz C, Puerta-Alcalde P, et al. Can artificial intelligence improve the management of pneumonia. J Clin Med 2020 Jan 17;9(1):248. [doi: 10.3390/jcm9010248] [Medline: 31963480]
32. Bradley JS, Byington CL, Shah SS, et al. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Clin Infect Dis 2011 Oct;53(7):e25-e76. [doi: 10.1093/cid/cir531] [Medline: 21880587]
33. Harris M, Clark J, Coote N, et al. British Thoracic Society guidelines for the management of community acquired pneumonia in children: update 2011. Thorax 2011 Oct;66 Suppl 2:ii1-i23. [doi: 10.1136/thoraxjnl-2011-200598] [Medline: 21903691]
34. Dean P, Florin TA. Factors associated with pneumonia severity in children: a systematic review. J Pediatric Infect Dis Soc 2018 Dec 3;7(4):323-334. [doi: 10.1093/jpids/piy046] [Medline: 29850828]
35. Araya S, Lovera D, Zarate C, et al. Application of a prognostic scale to estimate the mortality of children hospitalized with community-acquired pneumonia. Pediatr Infect Dis J 2016 Apr;35(4):369-373. [doi: 10.1097/INF.0000000000001018] [Medline: 26629871]
36. Williams DJ, Zhu Y, Grijalva CG, et al. Predicting severe pneumonia outcomes in children. Pediatrics 2016 Oct;138(4):e20161019. [doi: 10.1542/peds.2016-1019] [Medline: 27688362]
37. Zeng M, Zou B, Wei F, Liu X, Wang L. Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data. Presented at: 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS); May 28-29, 2016; Chongqing, China p. 225-228. [doi: 10.1109/ICOACS.2016.7563084]
38. Liu R, Greenstein JL, Fackler JC, Bergmann J, Bembea MM, Winslow RL. Prediction of impending septic shock in children with sepsis. Crit Care Explor 2021 Jun 15;3(6):e0442. [doi: 10.1097/CCE.0000000000000442] [Medline: 34151278]
39. Akhtar F, Li J, Pei Y, Xu Y, Rajput A, Wang Q. Optimal features subset selection for large for gestational age classification using GridSearch based recursive feature elimination with cross-validation scheme. In: Hung J, Yen N, Chang JW, editors. Frontier Computing: Theory, Technologies and Applications (FC 2019). Lecture Notes in Electrical Engineering, vol 551: Springer; 2020:63-71. [doi: 10.1007/978-981-15-3250-4_8]
40. Man X, Chan EP. The best way to select features? comparing MDA, LIME, and SHAP. J Financ Data Sci Winter 2021;3(1):127-139. [doi: 10.3905/jfds.2020.1.047]
41. GPT-4o. OpenAI. URL: https://platform.openai.com/docs/models/gpt-4o [accessed 2025-02-12]

Abbreviations

AUC-ROC: area under the receiver operator characteristic curve
CAP: community-acquired pneumonia
IMCI: Integrated Management of Childhood Illness
LightGBM: light gradient boosting machine
LMIC: low- and middle-income country
ML: machine learning
RFECV: recursive feature elimination with cross-validation
SHAP: Shapley additive explanations
SMOTE: Synthetic Minority Oversampling Technique
WHO: World Health Organization
XGBoost: extreme gradient boosting

Edited by S Amal; submitted 24.02.24; peer-reviewed by Anonymous, C Rogerson; revised version received 19.12.24; accepted 08.01.25; published 04.03.25.

Please cite as:
Serin O, Akbasli IT, Cetin SB, Koseoglu B, Deveci AF, Ugur MZ, Ozsurekci Y
Predicting Escalation of Care for Childhood Pneumonia Using Machine Learning: Retrospective Analysis and Model Development
JMIRx Med 2025;6:e57719
URL: https://xmed.jmir.org/2025/1/e57719
doi:10.2196/57719

Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study

Abdul Aziz Tayoun, MPH

School of Medicine, Department of Family and Community Medicine, Jordan University, Qween Rania Street, Amman, Jordan

Corresponding Author:
Abdul Aziz Tayoun, MPH
School of Medicine, Department of Family and Community Medicine, Jordan University, Qween Rania Street, Amman, Jordan

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.02.03.24302286v1

Companion article: https://med.jmirx.org/2025/1/e71529

Companion article: https://med.jmirx.org/2025/1/e71531

Companion article: https://med.jmirx.org/2025/1/e71528

Abstract

Background: Routine periodic health examinations (PHEs) for adults who are asymptomatic are included in clinical preventive services. They aim to prevent morbidity and mortality by identifying modifiable risk factors and early signs of treatable diseases. PHEs are a standard procedure in primary health care worldwide, including in Jordan. The country is undergoing an epidemiological transition toward noncommunicable diseases, which are the leading causes of morbidity and mortality. The prevalence of smoking is among the highest in the world, with escalating rates of obesity and physical inactivity. Notably, hypertension and diabetes are the most prevalent diseases.

Objective: This study aims to determine the extent to which individuals in Jordan participate in PHEs and to evaluate the various factors related to sociodemographics, health, knowledge, and behavior that influence this participation.

Methods: This study used a cross-sectional design and includes 362 participants 18 years or older residing in Jordan. A convenience sampling method was used, and data were collected through a hybrid web-based and face-to-face questionnaire. The analysis involved the application of logistic regression through SPSS to investigate the relationship between various influencing factors and the uptake of PHEs.

Results: Our study indicated that only 98 of the 362 (27.1%, 95% CI 22.8%-31.9%) participants underwent PHEs within the last 2 years. Noteworthy predictors of PHE uptake among Jordanians included recent visits to a primary health care facility within the previous year (adjusted odds ratio [AOR] 4.32, 95% CI 2.40 - 7.76; *P*<.001), monthly income (*P*=.02; individuals with a monthly income of 1500 - 2000 JD displayed more than five times the odds of undertaking PHEs than those with a monthly income <500 JD; AOR 5.74, 95% CI 1.32 - 24.90; *P*=.02; those with a monthly income of more than 2000 JD exhibited even higher odds; AOR 9.81, 95% CI 1.73 - 55.55; *P*=.02; a currency exchange rate of 1 JD=US $1.43 is applicable), and knowledge levels regarding PHEs and preventive health measures (AOR 1.23, 95% CI 1.03 - 1.47; *P*=.007). These variables emerged as the strongest predictors in our analysis, shedding light on key factors influencing PHE uptake in the population. Contrary to other research, our study did not find any statistically significant association between gender (*P*=.33), smoking status (*P*=.76), marital status (*P*=.52), health status self-evaluation (*P*=.18), seasonal influenza vaccination (*P*=.07), combined health behavior factors (*P*=.34), and BMI (*P*=.76) and PHE uptake.

Conclusions: PHE uptake is notably low in Jordan. Critical determinants of this uptake include recent visits to a primary health care facility within the previous year, monthly income, and knowledge levels regarding PHEs and preventive health services. To enhance PHE uptake, there is a critical need to integrate PHEs with primary health care services, increase awareness about PHEs, and offer free preventive services, particularly for those at high risk.

(JMIRx Med 2025;6:e57597) doi:10.2196/57597

KEYWORDS

periodic health examination; PHE; preventive health services; routine health checkups; Jordan; cross-sectional study

Introduction

Background

Routine periodic health examinations (PHEs) for adults who are asymptomatic are integral to primary health care practice. These examinations involve clinical preventive services administered by primary health care clinicians to individuals without signs or symptoms of illness, constituting a routine health care process. The goal of these examinations is to prevent morbidity and mortality proactively, this is achieved by identifying modifiable risk factors and detecting early signs of treatable diseases [1].

The health belief model (HBM) was conceptualized to elucidate why individuals are reluctant to engage in disease prevention programs and health checkups. As a crucial predictive framework, the HBM aids in understanding various health-related behaviors, including smoking, exercise, patient roles, and use of medical services [2].

Integrating with the HBM, health beliefs are defined as personal convictions associated with perceiving and managing specific diseases. These beliefs encompass key elements: perceived sensitivity, perceived severity, perceived benefit, perceived barrier, and cue to action [3].

Literature Review

A systematic review recently published in the *Canadian Family Physician Journal* aimed to assess the reasons for visits to primary health care clinics. Clinicians participating in the review identified routine health maintenance as the third most prevalent reason for individuals seeking consultations with primary health care physicians. This ranking positioned routine health maintenance after upper respiratory tract infections and hypertension, highlighting the significant role of primary health care practitioners in motivating individuals to engage with preventive health services [4].

A study conducted among undergraduate students in a Nigerian health science college found that 91.2% of participants demonstrated awareness of PHEs. However, the actual participation in PHEs was notably low at 28.4%. The primary obstacles to uptake were identified as insufficient time, religious considerations, duration of education, perceived susceptibility to diseases, financial constraints, apprehension about the results, and a general lack of interest [5].

A nationwide study in Saudi Arabia revealed that 22.9% of participants 15 years or older had undergone a PHE in the preceding 2 years. The probability of receiving a PHE during this period exhibited positive correlations with various factors—including age; educational attainment; marital status; regular consumption of five servings of fruits and vegetables daily; and diagnoses such as prediabetes, diabetes, or hypercholesterolemia—visit to a health care setting within the last 2 years due to illness or injury [6].

Rationale and Significance of the Study

Jordan, classified as an upper middle–income country, spans an area of 89,318 square kilometers and is divided into four provinces and 12 governorates. The population has grown substantially, increasing from 5.4 million in 2003 to over 11.5 million in 2023. This demographic shift can be attributed mainly to the influx of refugees and a relatively high birth rate [7,8].

The country is undergoing a notable epidemiological transition characterized by a rising prevalence of noncommunicable diseases (NCDs). These diseases are responsible for approximately 78% of deaths, establishing themselves as the primary cause of mortality and morbidity among the Jordanian population. Key risk factors contributing to the burden of NCDs include tobacco use, with a prevalence of about 50% (including e-cigarettes and shisha). One-quarter of the population reports insufficient physical activity and approximately 60% are classified as overweight or obese. Additionally, 22% of the population has hypertension, 14% has diabetes, and about 18% has depression [9].

Goals of This Study

This profile underscores a pressing concern regarding the country's high risk of NCDs. There is a need for evidence-based preventive health measures to curb the progression of NCDs and their associated risk factors. If conducted according to evidence-based guidelines, PHEs can effectively control communicable diseases and NCDs. Recognizing the urgency of the situation, gathering data on the uptake rate of PHEs, and identifying the factors influencing this uptake is imperative. The absence of previous studies on the uptake of PHEs in Jordan underscores the necessity for comprehensive research. Our study aims to estimate the uptake of PHEs among Jordanians while concurrently investigating various sociodemographic, health status, knowledge, and behavioral factors that play a role in influencing this uptake. The findings from this research will not only contribute valuable insights into the current scenario but also guide educational and promotional activities to encourage citizens to use preventive health services. In doing so, we strive to fill a crucial gap in existing knowledge and provide a foundation for evidence-based strategies to enhance public health in the country.

Methods

Recruitment

This descriptive cross-sectional study was conducted using an anonymous web-based Google Forms questionnaire between March 15 and May 1, 2023. Due to the lack of resources, a convenience sampling method was used to recruit participants. Jordanian residents aged ≥18 years who agreed to participate in our study were considered eligible. The research uses a questionnaire with five key domains: sociodemographic, health status, PHE uptake history, knowledge about PHEs, and health behaviors based on the HBM. This questionnaire was sent through the WhatsApp and Facebook platforms to participants, who were encouraged to share them with their family members. In addition, collecting data through face-to-face interviews targeted clients of grand malls, mosques, and pharmacies, supplementing the web-based data collection.

The study adopted a stratified proportional sampling strategy across four provinces of Jordan. This approach is carefully extended to maintain a balance in gender and nationality among

participants. The initial page of the web-based questionnaire explicitly outlines the study's objectives and provides detailed instructions on how to complete the questionnaire. This effort was complemented by the researcher's availability to answer questions, ensuring participants' queries or doubts were promptly addressed.

Sampling Method

The following inclusion and exclusion criteria were used:

- Inclusion criteria: any citizen regardless of nationality, 18 years or older, and residing in Jordan
- Exclusion criteria: persons younger than 18 years and individuals who declined to participate in the study

We recruited 362 respondents, aiming to provide a representative sample that reflects the entire population of Jordan in terms of district, age, sex, and nationality. The convenience sample size of 362 was calculated using the sample size formula for proportions:

$$N=Z\propto/22P1-PD2$$

This calculation considered a study conducted in Saudi Arabia, where approximately 34% of the population underwent PHEs [10]. The chosen values for statistical significance (α error) and margin of error (D) were .05% and 5%, respectively. As a result, the calculated sample size required for the survey was 345 respondents.

Questionnaire Development

The PHE questionnaire (Multimedia Appendix 1), comprising 36 questions across five domains, was developed following an extensive literature review [10-14]. The questionnaire's five domains are as follows:

1. Sociodemographic (9 items): inquires about relevant sociodemographic variables of participants
2. Health status and risk factors (7 items): explores participants' health status and associated risk factors
3. PHE uptake (4 items): focuses on the outcome variable of PHE uptake
4. Knowledge about PHE and preventive health services (8 items): assesses knowledge using a 3-option scale (agree, don't agree, I don't know). The items are scored, with correct answers receiving a score of 1 and incorrect or I don't know responses scoring 0. The total score ranges from 0 to 8, with higher scores indicating more significant knowledge of health checkups and preventive measures. The Cronbach α, estimated during the pilot phase with 25 participants, was 0.68.
5. Health behaviors toward PHE based on the HBM (6 items): measures health behaviors using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The total score ranges from 6 to 30, with higher scores indicating more positive health beliefs for each item. The Cronbach α for health behaviors toward PHEs during the pilot testing phase was 0.74, demonstrating acceptable internal consistency.

The questionnaire was translated into Arabic for comprehensibility and then back to English with the assistance of an expert translator. This rigorous process ensures the questionnaire's clarity and accuracy across languages.

Statistical Analysis

The primary outcome variable is the uptake of PHEs in Jordan, categorized as a dichotomous (yes or no) variable. The independent variables encompass sociodemographics, health status, knowledge, and health behavioral factors. Records with missing data were excluded to ensure the integrity of the analysis. Data was analyzed using SPSS, version 26.0 (IBM Corp).

Participant characteristics were examined using counts, percentages, means, and SDs through descriptive statistics. Graphs and tables were used as needed for visual representation. A 95% CI was calculated using appropriate methods, and a 2-sided *P* value <.05 was considered statistically significant.

A binary logistic regression test was used to study the association between the binary outcome variable and the various continuous and nominal predictor variables. Multivariate logistic regression analysis was used to examine the relationship between the uptake of PHEs and various independent covariables to adjust for confounding.

A hierarchical block-wise logistic regression model was also constructed to identify the most potent predictor variables. This comprehensive approach blends descriptive, inferential, and multivariate statistical techniques to provide a thorough understanding of the factors influencing the uptake of PHEs in Jordan.

Ethical Considerations

Before the formal survey, the study protocol was approved by the Jordan University Ethics Committee (approval 13 - 2023) and the Jordan University Hospital Ethics Committee (approval 10/2023/4560). The questionnaire was designed to be anonymous and voluntary, and respondents were informed that submission of the questionnaire implied informed consent. The data were kept confidential, and the results did not identify the respondents personally. Contact information for the researcher was provided for clarification purposes. No compensation was provided to participants.

Results

A total of 365 individuals participated in the study between March and April 2023, with a response rate of 99%; 3 participants were excluded (one was younger than 18 years, and the other two did not complete the questionnaire), leaving 362 participants for analysis.

Descriptive Statistics

The demographic characteristics of participants are summarized in Table 1. The mean age was 38.2 (SD 14.6, range 18-88) years. Of the 362 participants, there were slightly more male (n=185, 51.1%) than female participants. Approximately 230 (63.5%) were married, 270 (74.6%) were Jordanians, and 202 (55.8%) held a university degree. Most participants (n=225, 62.1%) reported a monthly income of less than 500 JD (a currency

exchange rate of 1 JD=US $1.43 is applicable), with half lacking health insurance.

Regarding health status, Table 2 shows that 240 (66.3%) participants reported good or excellent health, 78 (21.5%) had a chronic disease, and 200 (55.2%) visited a primary health care clinic in the past year. Additionally, 191 (52.8%) participants were current smokers.

Regarding PHEs, only 98 of the 362 (27.1%, 95% CI 22.8% - 31.9%) participants underwent a medical checkup in the last 2 years.

Table . Sociodemographic characteristics of participants (N=362).

Characteristic	Participants, n (%)
Gender	
Male	185 (51.1)
Age group (years)	
18 - 29	122 (33.7)
30 - 39	90 (24.9)
40 - 49	70 (19.3)
50 - 59	41 (11.3)
≥60	39 (10.8)
Marital status	
Married	230 (63.5)
Single	101 (27.9)
Divorced	14 (3.9)
Widowed	17 (4.7)
Monthly income (JD)[a]	
<500	225 (62.1)
500 - 999	93 (25.7)
1000 - 1499	26 (7.2)
1500 - 1999	10 (2.8)
≥2000	8 (2.2)
Educational level	
Elementary school	42 (11.6)
Secondary school	118 (32.6)
University	166 (45.9)
Postgraduate	36 (9.9)
Province of residence	
Amman	151 (41.7)
Central Jordan	82 (22.7)
North Jordan	100 (27.6)
South Jordan	29 (8.0)
Nationality	
Jordanians	270 (74.6)
Syrians	47 (13.0)
Palestinians	22 (6.1)
Egyptians	18 (5.0)
Iraqis	5 (1.4)

[a]A currency exchange rate of 1 JD=US $1.43 is applicable.

Table . Health characteristics of participants in the study.

Variable	Participants, n (%)
Visiting a primary health care facility within the previous year	
Yes	200 (55.2)
No	162 (44.8)
Noncommunicable diseases	
Yes	78 (21.5)
No	284 (78.5)
Smoking	
Smoker	191 (52.8)
Not smoker	171 (47.2)
Health insurance	
Insured	183 (50.6)
Not insured	179 (49.4)
Seasonal flu vaccination	
Yes	60 (16.6)
No	302 (83.4)
Health status self-evaluation	
Poor	9 (2.5)
Fair	25 (6.9)
Good	88 (24.3)
Very good	136 (37.6)
Excellent	104 (28.7)
BMI≥25	
Yes	223 (61.6)
No	139 (38.4)
Physical activity	
Yes	108 (29.8)
No	254 (70.2)

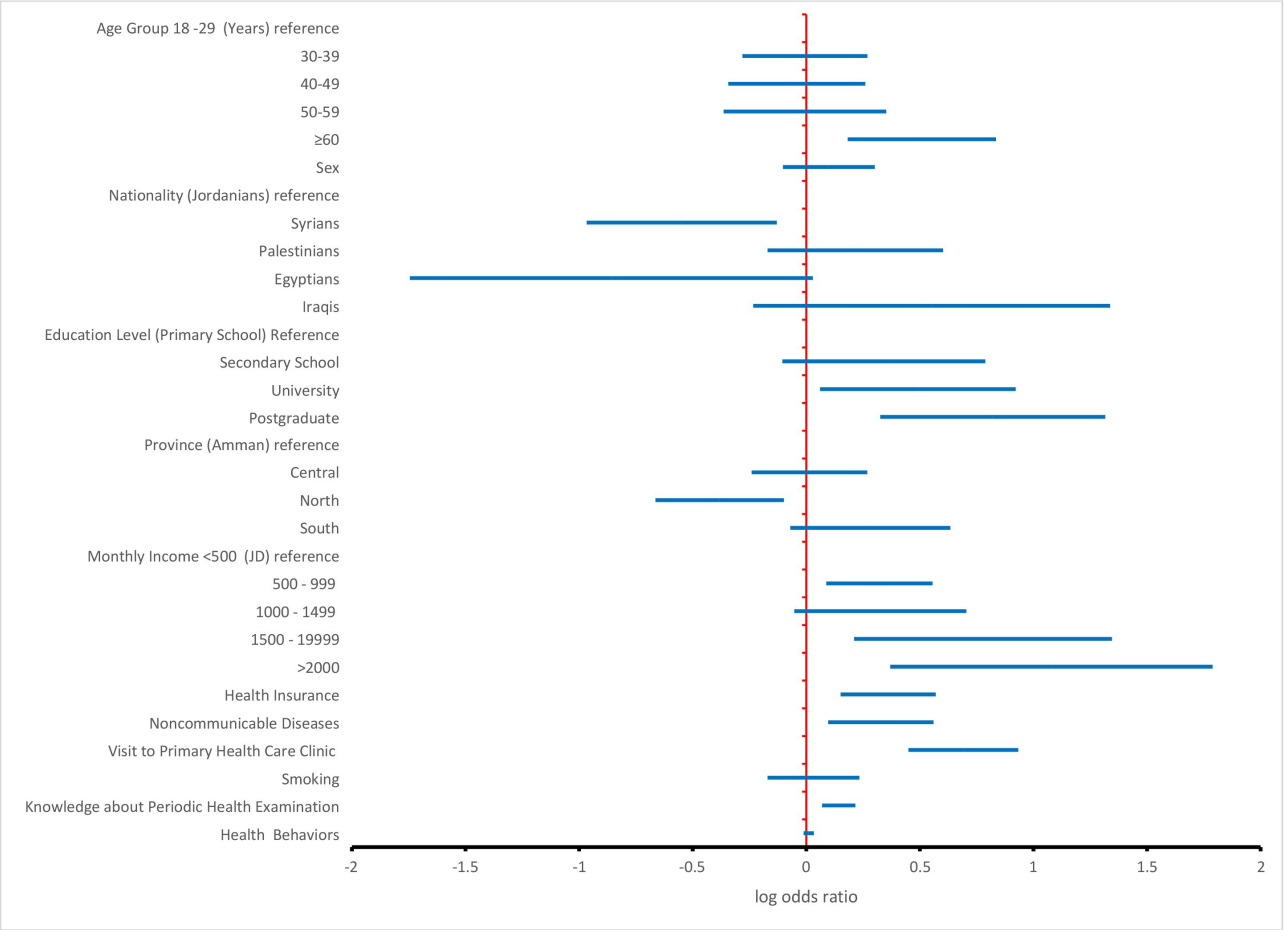
Logistic Regression Analysis

The forest plot in Figure 1 highlights several significant findings from the analysis of the predicting factors' association with PHE uptake.

Age was found to be a significant determinant of PHE uptake: with each additional year of age there, is a 2.2% increase in the odds of undertaking PHEs (odds ratio [OR] 1.022, 95% CI 1.006 - 1.038; *P*=.006). Nationality also proved to be a factor, with Syrians demonstrating a lower frequency of PHE uptake. The odds of Syrians undergoing PHEs were 0.283 compared to Jordanians (OR 0.28, 95% CI 0.11 - 0.74; *P*=.01). Education level exhibited a strong association, with postgraduates displaying more than 6 times the odds of undertaking PHE than individuals with only primary school education (OR 6.62, 95% CI 2.12 - 20.71; *P*=.001). Health care workers displayed more than 12 times the odds of undergoing PHEs than general employees (OR 12.28, 95% CI 4.69 - 32.19; *P*<.001). Individuals earning more than 2000 JD monthly had 12 times greater odds of receiving PHEs compared to those with a monthly income of less than 500 JD (OR 12.00, 95% CI 2.34 - 61.45; *P*=.003). Health insurance emerged as a significant facilitator of PHE uptake. Insured participants demonstrated more than 2 times the odds of undertaking PHEs than noninsured individuals (OR 2.30, 95% CI 1.42 - 3.71; *P*=.001). People with chronic diseases have more than twice the odds of undertaking PHEs than those without chronic diseases (OR 2.3, 95% CI 1.258 - 3.629; *P*=.005). Visits to a primary health care clinic in the past year significantly impacted PHE uptake. Those who had visited had 5 times the odds of PHE uptake compared to those who did not visit a primary health care facility in the past year (OR 4.91, 95% CI 2.82 - 8.57; *P*<.001). Participants who were physically active had 1.65 times the odds of undertaking PHEs than those without enough physical activity (OR 1.65, 95% CI 1.01-2.69; *P*=.046). Finally, for every extra point in knowledge about PHEs, there is a 39% increase in PHE uptake (OR 1.39, 95% CI 1.18 - 1.64; *P*<.001).

On the other hand, several variables were not associated with PHE uptake. These included gender (*P*=.33), smoking status (*P*=.76), marital status (*P*=.52), health status self-evaluation (*P*=.18), seasonal influenza vaccination (*P*=.07), combined health behavior factors (*P*=.34), and BMI (*P*=.76).

Figure 1. Univariate logistic regression analysis for predictor factors of periodic health examination uptake, Jordan 2023. A currency exchange rate of 1 JD=US $1.43 is applicable.



Adjusted Logistic Regression Model

After meticulously adjusting for confounding variables and carefully selecting clinically and statistically significant factors, we successfully constructed a logistic regression model using the hierarchical block-wise method. This refined model, depicted in Table 3, encapsulates three variables that significantly influence the uptake of PHEs.

Table . Logistic regression model for most significant predictor factors for periodic health examination uptake, Jordan 2023.

Variable	*P* value	Adjusted odds ratio (95% CI)
Visiting a primary health care facility	<.001	4.315 (2.40-7.76)
Knowledge about periodic health examinations	.02	1.230 (1.03-1.47)
Monthly income (JD)[a]	.07	
<500 (reference)	—[b]	1.00
500 - 999	.07	1.71 (0.96-3.02)
1000 - 1499	.11	2.18 (0.84-5.66)
1500 - 1999	.02	5.74 (1.32-24.90)
≥2000	.01	9.81 (1.73-55.55)

[a]A currency exchange rate of 1 JD=US $1.43 is applicable.

[b]Not applicable.

Visit to Primary Health Care Facilities in the Past Year

Visiting primary health care facilities within the past year exhibited a substantial impact on PHE uptake. These individuals demonstrated more than 4 times the odds of undertaking PHEs compared to those who did not visit a primary health care facility within the same time frame (adjusted OR [AOR] 4.32, 95% CI 2.40 - 7.76; *P*<.001).

Income Level

Individuals with a monthly income of 1500 - 2000 JD displayed more than five times the odds of undertaking PHEs than those with a monthly income of less than 500 JD (AOR 5.74, 95% CI 1.32 - 24.90; *P*=.02). Furthermore, those with a monthly income of more than 2000 JD exhibited even higher odds (AOR 9.81, 95% CI 1.73 - 55.55; *P*=.02).

Health Knowledge

The analysis indicates that for every point increase in PHE knowledge, the likelihood of individuals opting for PHEs increases by 23% (AOR 1.23, 95% CI 1.03-1.47; *P*=.02).

Discussion

Principal Findings and Comparison With Other Studies

Of the 362 participants, only 98 (27.1%, 95% CI 22.8%-31.9%) had undergone a PHE in the past 2 years. Some significant predictors included recent visits to a primary health care facility the previous year, monthly income, knowledge about PHEs, and preventive health measures. Other nonsignificant factors were gender, marital status, smoking status, and BMI, which did not emerge as being significantly associated with the uptake of PHEs.

Interestingly, the uptake rate observed in our study is comparable to that reported in studies conducted in Saudi Arabia [6,10] and Nigeria [12]. In contrast, this rate notably fell below those reported in studies conducted in the United States [1], the United Kingdom [13], and Switzerland [15].

The most influential determinant for the uptake of PHEs found in our study was a visit to a primary health care facility in the past year. Our findings again were consistent with those from several other studies [6,16,17]. Notably, those who had visited any primary health care clinic in the previous year were found to be five times more likely to undertake PHEs compared to those who had not visited such clinics in the same time frame. This association was statistically significant even after adjusting for other relevant factors, thus underlining its strength. The second most important factor influencing the uptake of PHEs was monthly income. This finding agrees with results from other sources [1,12,14,17-21]. The influence of monthly income on the uptake of PHEs reflects how socioeconomic issues can affect health care–seeking behavior. There is a great need for focused efforts or an intervention policy that addresses these issues. Knowledge about PHEs was the third most influential factor. The findings are in agreement with those of previous studies [22-24] and underline the role of informed choice in health care use. This paper should, however, state that knowledge of PHEs was associated with other factors such as educational level and occupation. However, adjustment for these factors associated with knowledge of PHEs did not reduce the strength of the association with knowledge and PHE uptake.

More variables were positively associated with the uptake of PHEs. The older the age, the better the PHE uptake, which agrees with other studies' findings [13,17,19]. This may indicate that with increased age, people are likely to undergo regular health checkups, either because of the higher burden of NCDs in older age or maybe because more attention is paid to preventive measures with increased age. Individuals of Syrian nationality were found to be less likely to undergo PHEs than Jordanians. Economic factors may explain this difference, emphasizing the need for targeted interventions to ensure equitable access to preventive health care services among diverse populations. There was a strong association between education and PHE uptake, evidenced by a substantial increase in PHE uptake corresponding to higher levels of education. This finding is similar to results from other studies [17,21,25]. Compared to employees in general, health workers and retirees were more likely to undergo PHEs. This may be because health care workers are more aware of the importance of preventive health. Age can serve as a confounder for retired people because it may affect retired status and PHE uptake.

The health-related factors identified to be associated with PHE uptake in our study, and supported by other studies, include the presence of chronic diseases [6,14,18,22,26], being insured [17,21,25,27,28], and engagement in physical activity [1].

Other factors showed no significant association with the uptake of PHEs. For example, one nonsignificant factor was sex, which contrasts many studies indicating that females are more willing to participate in PHEs than males [6,13,15,20]. Being married has often been linked to higher PHE uptake in previous studies but not in our study [1,13-15,19,29,30]. Surprisingly, smoking status was not associated with the uptake of PHEs; several studies in the past have argued that smokers are less likely than nonsmokers to undergo PHEs [11,13,15,20,29]. Our study did not find any clear association between combined behavioral factors and the uptake of PHEs, although many studies identify such associations [3,11,14,20,30,31]. This is possibly because of the suitability of the questionnaires to the Jordanian population or problems with participants understanding.

Strengths of the Study

This study is the first of its kind to investigate the uptake of PHEs in Jordan and hence addresses an important gap in existing knowledge. Given that this is the first study on this topic, it has contributed quite substantially to the understanding of PHE uptake in the country. The statistical analysis approach adopted in this study is broad and solid, using descriptive, inference, and multivariate statistical techniques. This approach leads to a deeper analysis and more reliable findings. The study also managed to identify the significant predictors of PHE uptake.

Limitations of the Study

One of the primary limitations is its cross-sectional design, which restricts the ability to establish causality between the different predictor factors and PHE uptake. To address this issue, future research could adopt a longitudinal approach,

providing a better understanding of how these predictors influence PHE uptake. Another limitation relates to the sampling method. The study used a convenience sampling strategy, which may have introduced selection bias, and the web-based survey format could lead to measurement bias. To decrease the chances of bias, we used a stratified sampling method, taking into account population size and stratifying participants by gender, age group, and nationality across the four provinces of Jordan. Additionally, a hybrid approach integrating both web-based and face-to-face interviews, and collecting data from various settings such as social media platforms, grand malls, mosques, and pharmacies helped ensure a more representative sample. The author's availability for clarifications via WhatsApp and email also aimed to reduce potential measurement biases during data collection. The third limitation concerns the survey instrument itself. The comprehensiveness and relevance of the questionnaire to the Jordanian population might not have been fully ensured. To address this issue, a pilot study with 25 participants was conducted, and the questionnaire was revised based on their feedback and reliability measures. Lastly, the study's results may have limited generalizability beyond the population of Jordan. To enhance the applicability of the findings to broader populations, future research should consider a more diverse sample by including other countries. This would provide a more comprehensive understanding of PHE uptake within and outside Jordan.

Future Directions

First, we established that recent visits to primary health care facilities were the strongest predictor of PHE uptake. From this, we recommend incorporating preventive health services into existing primary health care services to enhance accessibility and efficiency. This may take the form of incentivizing both health caregivers and patients. Second, economic issues can be resolved by suggesting the provision of all preventive services free of cost at primary health care centers. Private health insurance companies can also facilitate this endeavor by covering preventive services like PHEs within the realm of their service provision so that people can have better access to these services. More importantly, public awareness will have to increase. The positive correlation between knowledge of PHEs and their uptake points to a need for more organized and evidence-based awareness campaigns. Another issue involves the study's findings on behavioral factors. The study did not find a significant relationship between behavioral factors and PHE uptake, contradicting findings from other contexts. To better understand these results, future research could involve a more detailed investigation into the cultural and societal influences on health behaviors in Jordan, which may help clarify why these factors did not show the expected association. It is also recommended that further studies, especially on smoking as a predictor factor for PHE uptake, be done in detail to understand how to best address these areas in future studies.

Conclusion

Our study has highlighted the low level of PHE uptake in Jordan. This paper identified visitation to primary health care facilities in the past year, monthly income, and knowledge about PHEs and preventive health services as the major predictors influencing the likelihood of undergoing PHEs. The association of regular visits to primary health care facilities with higher uptake of PHEs suggests that PHEs should be integrated with the available services at primary health care facilities. These findings also suggest that targeted interventions should be implemented to enhance awareness and knowledge of the value of preventive health practices among the Jordanian population, particularly for patients with lower income status.

Acknowledgments

We conducted this review using the ChatGPT-4 model, 2023, developed by OpenAI, which only helped create text. The author reviewed and edited ChatGPT's draft for accuracy and coherence (Multimedia Appendix 2). We are grateful to all the survey participants, and the Jordan University and Jordan University Hospital ethical committees.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

AAT analyzed the data, drafted the manuscript, and devised the study concept and design. Furthermore, AAT interpreted the results and is responsible for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire in English.
[DOCX File, 20 KB - xmed_v6i1e57597_app1.docx]

Multimedia Appendix 2
ChatGPT's draft.

[DOCX File, 32 KB - xmed_v6i1e57597_app2.docx]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[DOCX File, 34 KB - xmed_v6i1e57597_app3.docx]

References

1. Culica D, Rohrer J, Ward M, Hilsenrath P, Pomrehn P. Medical checkups: who does not get them? Am J Public Health 2002 Jan;92(1):88-91. [doi: 10.2105/ajph.92.1.88] [Medline: 11772768]
2. Cho MK, Cho YH. Role of perception, health beliefs, and health knowledge in intentions to receive health checkups among young adults in Korea. Int J Environ Res Public Health 2022 Oct 24;19(21):13820. [doi: 10.3390/ijerph192113820] [Medline: 36360698]
3. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [Medline: 7738325]
4. Finley CR, Chan DS, Garrison S, et al. What are the most common conditions in primary care? Systematic review. Can Fam Physician 2018 Nov;64(11):832-840. [Medline: 30429181]
5. Esan O, Akinyemi A, Ayegbusi O, Bakare T, Balogun Y, Ogunwusi A. Determinants of uptake of periodic medical examination among students of college of health sciences, Obafemi Awolowo University Ile-Ife, South-West Nigeria. Nigerian J Med 2020;29(4):575-581. [doi: 10.4103/NJM.NJM_150_20]
6. El Bcheraoui C, Tuffaha M, Daoud F, et al. Low uptake of periodic health examinations in the Kingdom of Saudi Arabia, 2013. J Family Med Prim Care 2015;4(3):342-346. [doi: 10.4103/2249-4863.161313] [Medline: 26288771]
7. Department of Statistics Jordan. URL: https://dosweb.dos.gov.jo/ [accessed 2023-01-14]
8. Jordan Ministry of Health. URL: https://moh.gov.jo/Default/En [accessed 2023-01-14]
9. Country cooperation strategy for WHO and Jordan 2021–2025: CCS Jordan. World Health Organization. 2021 Jun 21. URL: https://www.who.int/publications/i/item/9789290227014 [accessed 2023-01-14]
10. Al-Kahil AB, Khawaja RA, Kadri AY, Abbarh Mbbs SM, Alakhras JT, Jaganathan PP. Knowledge and practices toward routine medical checkup among middle-aged and elderly people of Riyadh. J Patient Exp 2020 Dec;7(6):1310-1315. [doi: 10.1177/2374373519851003] [Medline: 33457580]
11. Zhang Z, Yin AT, Bian Y. Willingness to receive periodic health examination based on the health belief model among the elderly in rural China: a cross-sectional study. Patient Prefer Adherence 2021 Jun 21;15:1347-1358. [doi: 10.2147/PPA.S312806] [Medline: 34188452]
12. Ofoli JNT, Ashau-Oladipo T, Hati SS, Ati L, Ede V. Preventive healthcare uptake in private hospitals in Nigeria: a cross-sectional survey (Nisa Premier Hospital). BMC Health Serv Res 2020 Apr 1;20(1):273. [doi: 10.1186/s12913-020-05117-5] [Medline: 32238153]
13. Labeit A, Peinemann F, Baker R. Utilisation of preventative health check-ups in the UK: findings from individual-level repeated cross-sectional data from 1992 to 2008. BMJ Open 2013 Dec 23;3(12):e003387. [doi: 10.1136/bmjopen-2013-003387] [Medline: 24366576]
14. Liu X, Li N, Liu C, et al. Urban-rural disparity in utilization of preventive care services in China. Medicine (Baltimore) 2016 Sep;95(37):e4783. [doi: 10.1097/MD.0000000000004783] [Medline: 27631229]
15. Diaz Hernandez L, Giezendanner S, Fischer R, Zeller A. Expectations about check-up examinations among Swiss residents: a nationwide population-based cross-sectional survey. PLoS One 2021 Jul 21;16(7):e0254700. [doi: 10.1371/journal.pone.0254700] [Medline: 34288961]
16. Alzahrani AMA, Felix HC, Stewart MK, Selig JP, Swindle T, Abdeldayem M. Utilization of routine medical checkup and factors influencing use of routine medical checkup among Saudi students studying in the USA in 2019. Saudi J Health Syst Res 2021 Mar 11;1(1):16-25. [doi: 10.1159/000514178]
17. Okoli GN, Abou-Setta AM, Neilson CJ, Chit A, Thommes E, Mahmud SM. Determinants of seasonal influenza vaccine uptake among the elderly in the United States: a systematic review and meta-analysis. Gerontol Geriatr Med 2019 Aug 17;5:2333721419870345. [doi: 10.1177/2333721419870345] [Medline: 31453267]
18. Getahun GK, Arega M, Keleb G, Shiferaw A, Bezabih D. Assessment of routine medical checkups for common noncommunicable diseases and associated factors among healthcare professionals in Addis Ababa, Ethiopia, in 2022 a cross-sectional study. Ann Med Surg (Lond) 2023 Apr 1;85(5):1633-1641. [doi: 10.1097/MS9.0000000000000558] [Medline: 37229001]
19. Dryden R, Williams B, McCowan C, Themessl-Huber M. What do we know about who does and does not attend general health checks? Findings from a narrative scoping review. BMC Public Health 2012 Aug 31;12:723. [doi: 10.1186/1471-2458-12-723] [Medline: 22938046]
20. Bjerregaard AL, Maindal HT, Bruun NH, Sandbæk A. Patterns of attendance to health checks in a municipality setting: the Danish “Check Your Health Preventive Program”. Prev Med Rep 2016 Dec 21;5:175-182. [doi: 10.1016/j.pmedr.2016.12.011] [Medline: 28050340]

21. Obi IR, Obi KM, Seer-Uke EN, Onuorah SI, Okafor NP. Preventive health care services utilization and its associated factors among older adults in rural communities in Anambra State, Nigeria. Pan Afr Med J 2021 May 28;39:83. [doi: 10.11604/pamj.2021.39.83.26997] [Medline: 34466185]
22. Alzahrani AM, Quronfulah BS, Felix HC, Khogeer AA. Barriers to routine checkups use among Saudis from the perspective of primary care providers: a qualitative study. Saudi Med J 2022 Jun;43(6):618-625. [doi: 10.15537/smj.2022.43.6.20220090] [Medline: 35675932]
23. Laz TH, Rahman M, Berenson AB. An update on human papillomavirus vaccine uptake among 11-17 year old girls in the United States: National Health Interview Survey, 2010. Vaccine (Auckl) 2012 May 21;30(24):3534-3540. [doi: 10.1016/j.vaccine.2012.03.067] [Medline: 22480927]
24. Sommer I, Titscher V, Gartlehner G. Participants' expectations and experiences with periodic health examinations in Austria - a qualitative study. BMC Health Serv Res 2018 Oct 30;18(1):823. [doi: 10.1186/s12913-018-3640-6] [Medline: 30376830]
25. AshaRani PV, Devi F, Wang P, et al. Factors influencing uptake of diabetes health screening: a mixed methods study in Asian population. BMC Public Health 2022 Aug 9;22(1):1511. [doi: 10.1186/s12889-022-13914-2] [Medline: 35941579]
26. Gosadi IM, Ayoub RA, Albrahim HT, et al. An assessment of the knowledge and practices of adults in Jazan, Saudi Arabia, concerning routine medical checkups. Patient Prefer Adherence 2022 Aug 5;16:1955-1969. [doi: 10.2147/PPA.S376345] [Medline: 35958888]
27. Leung JKF, Wong MCS, Wong ELY. Unseen threats of chronic diseases among the middle-aged: examining the feasibility of well-defined healthcare vouchers in encouraging uptake of general checkups. Int J Environ Res Public Health 2022 Sep 17;19(18):18. [doi: 10.3390/ijerph191811751] [Medline: 36142023]
28. Cherrington A, Corbie-Smith G, Pathman DE. Do adults who believe in periodic health examinations receive more clinical preventive services? Prev Med 2007 Oct;45(4):282-289. [doi: 10.1016/j.ypmed.2007.05.016] [Medline: 17692368]
29. Mori Y, Matsushita K, Inoue K, Fukuma S. Patterns and predictors of adherence to follow-up health guidance invitations in a general health check-up program in Japan: a cohort study with an employer-sponsored insurer database. PLoS One 2023 May 25;18(5):e0286317. [doi: 10.1371/journal.pone.0286317] [Medline: 37228080]
30. Zhang J, Oldenburg B, Turrell G. Measuring factors that influence the utilisation of preventive care services provided by general practitioners in Australia. BMC Health Serv Res 2009 Dec 3;9:218. [doi: 10.1186/1472-6963-9-218] [Medline: 19954549]
31. Oboler SK, Prochazka AV, Gonzales R, Xu S, Anderson RJ. Public expectations and attitudes for annual physical examinations and testing. Ann Intern Med 2002 May 7;136(9):652-659. [doi: 10.7326/0003-4819-136-9-200205070-00007] [Medline: 11992300]

Abbreviations

AOR: adjusted odds ratio
HBM: health belief model
NCD: noncommunicable disease
OR: odds ratio
PHE: periodic health examination

Edited by T Leung; submitted 20.02.24; peer-reviewed by A Ahmed, Anonymous; revised version received 17.10.24; accepted 26.10.24; published 05.02.25.

Please cite as:
Tayoun AA
Determinants of Periodic Health Examination Uptake: Insights From a Jordanian Cross-Sectional Study
JMIRx Med 2025;6:e57597
URL: https://xmed.jmir.org/2025/1/e57597
doi:10.2196/57597

Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis

Hojjat Borhany, MSc

Faculty of Environmental Science, Department of Environmental Science, Informatic, and Statistics, University of Ca' Foscari Venice, Mestre (VE), Italy

Corresponding Author:
Hojjat Borhany, MSc
Faculty of Environmental Science, Department of Environmental Science, Informatic, and Statistics, University of Ca' Foscari Venice, Mestre (VE), Italy

Related Articles:

Companion article: https://www.biorxiv.org/content/10.1101/2023.06.21.545938v2

Companion article: https://med.jmirx.org/2025/1/e69895

Companion article: https://med.jmirx.org/2025/1/e69896

Companion article: https://med.jmirx.org/2025/1/e69894

Abstract

Background: Italy can augment its profit from biorefinery products by altering the operation of digesters or different designs to obtain more precious bioproducts like volatile fatty acids (VFAs) than biogas from organic municipal solid waste. In this context, recognizing the process stability and outputs through operational interventions and its technical and economic feasibility is a critical issue. Hence, this study involves an anaerobic digester in Treviso in northern Italy.

Objective: This research compares a novel line, consisting of pretreatment, acidogenic fermentation, and anaerobic digestion, with single-step anaerobic digestion regarding financial profit and surplus energy. Therefore, a mass flow model was created and refined based on the outputs from the experimental and numerical studies. These studies examine the influence of hydraulic retention time (HRT), pretreatment, biochar addition, and fine-tuned feedstock/inoculum (FS/IN) ratio on bioproducts and operational parameters.

Methods: VFA concentration, VFA weight ratio distribution, and biogas yield were quantified by gas chromatography. A *t* test was then conducted to analyze the significance of dissimilar HRTs in changing the VFA content. Further, a feasible biochar dosage was identified for an assumed FS/IN ratio with an adequately long HRT using the first-order rate model. Accordingly, the parameters for a mass flow model were adopted for 70,000 population equivalents to determine the payback period and surplus energy for two scenarios. We also explored the effectiveness of amendments in improving the process kinetics.

Results: Both HRTs were identical concerning the ratio of VFA/soluble chemical oxygen demand (0.88 kg/kg) and VFA weight ratio distribution: mainly, acetic acid (40%), butyric acid (24%), and caproic acid (17%). However, a significantly higher mean VFA content was confirmed for an HRT of 4.5 days than the quantity for an HRT of 3 days (30.77, SD 2.82 vs 27.66, SD 2.45 g–soluble chemical oxygen demand/L), using a *t* test (t_8=−2.68; *P*=.03; CI=95%). In this research, 83% of the fermented volatile solids were converted into biogas to obtain a specific methane (CH_4) production of 0.133 CH_4-Nm^3/kg–volatile solids. While biochar addition improved only the maximum methane content by 20% (86% volumetric basis [v/v]), the FS/IN ratio of 0.3 volatile solid basis with thermal plus fermentative pretreatment improved the hydrolysis rate substantially (0.57 vs 0.07, 1/d). Furthermore, the biochar dosage of 0.12 g-biochar/g–volatile solids with an HRT of 20 days was identified as a feasible solution. Principally, the payback period for our novel line would be almost 2 years with surplus energy of 2251 megajoules [MJ] per day compared to 45 years and 21,567 MJ per day for single-step anaerobic digestion.

Conclusions: This research elaborates on the advantage of the refined novel line over the single-step anaerobic digestion and confirms its financial and technical feasibility. Further, changing the HRT and other amendments significantly raised the VFA concentration and the process kinetics and stability.

(JMIRx Med 2025;6:e50458) doi:10.2196/50458

KEYWORDS

multistep fermentation; specific methane production; anaerobic digestion; kinetics study; biochar; first-order; modified Gompertz; mass balance; waste management; environment sustainability

Introduction

The European Union annually generated about 110 million tons of organic waste in 2006, which excluded slurry and manure. This waste mainly came from the food industry (33%), agriculture and hunting (30%), and households (20%) [1]. Current Italian legislation forbids landfilling organic waste and requires treating it through biological and thermal processes like anaerobic digestion, composting, and incineration with high disposal costs for secondary waste flux (€75 - €125 per ton; a currency exchange rate of €1=US $1.05 is applicable) [2]. Under the pressure of exhaustible natural exploitation and increasing organic waste, the European Commission approved the circular economy action plan to promote sustainable recovery methods to reduce the secondary waste flux. The techniques recommended in the circular economy context assume a "take-use-reuse" viewpoint. Such an approach wants to close the circuit of cycles, extend product life, and treat the wastes as precious recyclable materials [3,4]. In this respect, the European Union states have deployed biological processes such as anaerobic digestion to gain either platform chemicals like volatile fatty acids (VFAs) or biogas from organic wastes produced in urban areas [5-9]. These products are extremely valuable in the era of environmental disasters, which have several consequences (eg, climate change), since they are renewable, sustainable, carbon-neutral, and compatible with current fossil-based fuel infrastructures [10].

Recent studies have aimed at finding a sequential reclaiming route to obtain various bioproducts such as VFAs and biohydrogen with a higher added-value market than bio-methane at distinct steps to either redesign the existing plants or integrate them into biorefinery platforms [11,12]. Various biological processes can convert different feedstock (eg, edible sugary crops, oil-bearing crops, livestock, waste sludge [WS], and food waste) into a range of biofuels, including bioethanol, biodiesel, bio-methane, and biohydrogen [10,13,14]. Biofuel production from edible crops is quite controversial in terms of food supply, ethical quandary, and insecure supply chain. However, food waste, WS, and livestock are omnipresent in urban and rural areas without widespread deployment in a biorefinery scheme. Accordingly, this research aims to convert organic municipal solid waste (OMSW), mainly from food waste, into VFAs and biogas.

This study examines the biological recovery route for OMSW for potential beneficial bioproducts and technical feasibility. This effort includes three steps: pretreatment, mesophilic acidogenic fermentation, and anaerobic digestion. Specifically, we endeavor to conceive how to make the process more profitable and practicable through operational amendments that change the share of methanogenesis and acidogenic routes in the final products (VFAs and biogas) [9] and lower the costs of the process in terms of energy and water consumption. Hence, determining a reasonably priced process with a desirable VFA-rich stream from acidogenic fermentation and a high methane (CH_4) yield from methanogenesis [15] could ultimately encourage full-scale commercialization. VFAs typically serve as platform chemicals for many processes (eg, biopolymer synthesis of polyhydroxyalkanoates [PHAs] [16-19]), which could be later recovered through biological processes to close the material life cycle.

The major bottleneck in anaerobic digestion of biowaste is at the hydrolysis step. Such a problem could be relieved by various methods such as pretreatment, an optimized feedstock/inoculum (FS/IN) ratio, and carbonaceous material addition, including biochar [20-22]. The latter method was recently realized to have numerous benefits to the process, such as improving the process stability, acceleration of the process rate, buffering potency and alkalinity, inhibitors adsorption, enriched microbial functionality, and electron transfer mechanism. As a result, it could improve CH_4 generation by fostering hydrolysis, acetogenesis, and methanogenesis [23]. The residual solids out of the multistep line of pretreatment followed by acidogenic fermentation plus anaerobic digestion can be used in a pyrolysis line for biochar and biofuel production to further lower the secondary waste flux [24]. This strategy provides several benefits, such as combating climate change and global soil degradation and addressing the rising energy demand.

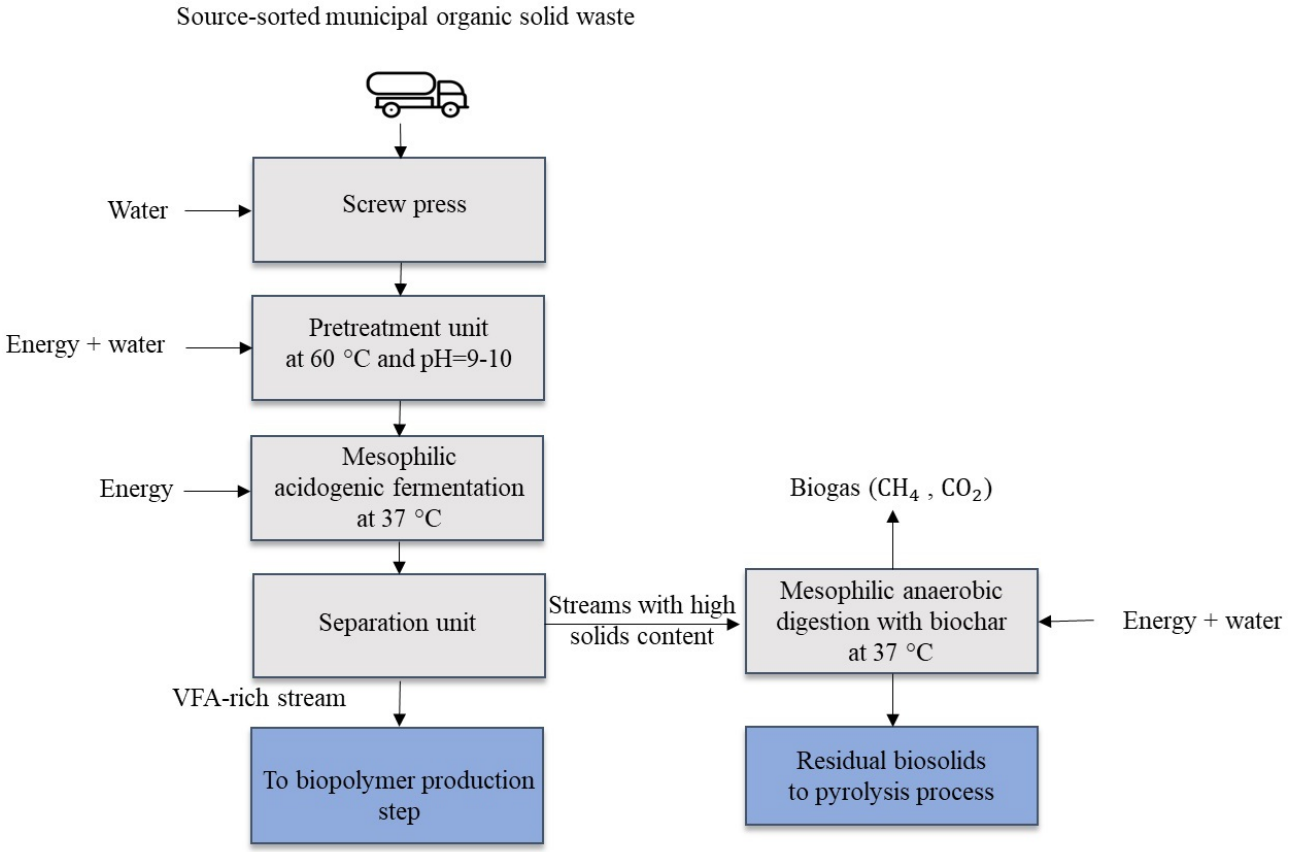
This study compares the multistep route of pretreatment, acidogenic fermentation, and anaerobic digestion with the existing method of single-step anaerobic digestion for valorizing OMSW in the Treviso wastewater treatment plant (WWTP) in terms of financial profit and technical feasibility. In this context, the present research has the ultimate goals of facilitating the entrance of the process into the market and further closure of the cycle of organic material. Accordingly, it assesses several suggestions, such as hydraulic retention time (HRT) variation, pretreatment, biochar addition, and adjusted FS/IN ratio to enhance the bioproducts and decrease the involved costs. To this end, their effects on the process were quantified through experimental tests, confirming their significance through statistical analysis. Later, the payback period, amount of surplus energy, and volatile solids (VS) destruction for the mentioned scenarios were determined using a mass balance model refined according to the laboratory studies. The boundary condition parameters for energy conversion and costs were assumed according to previous studies and experts' knowledge, respectively. To the best of the author's knowledge, this paper is novel in presenting a robust framework to assess a groundbreaking proposition for the valorization of OMSW financially and technically. Overall, we concluded that our line is viable technically and overtakes the conventional methods financially.

Methods

Biorefinery Process Scheme and Experimental Studies

Figure 1 presents the hypothesized biorefinery process line in this research. It comprises screw-pressing, a pretreatment unit, mesophilic acidogenic fermentation, solid-liquid separator, and mesophilic anaerobic digestion. The two sectors of biopolymer production and pyrolysis were exhibited differently since no mass and energy flow was considered for them, and only the possible end goals for the secondary stream were shown.

Figure 1. Schematic of the multistep process of pretreatment, acidogenic fermentation, and anaerobic digestion for VFAs and biogas production from the organic municipal solid waste. CH_4: methane; CO_2: carbon dioxide; VFA: volatile fatty acid.



After and before the pretreatment, the feedstock for different parameters was characterized from time to time. These parameters include the total solids (TS), VS, chemical oxygen demand (COD), soluble COD (SCOD), total Kjeldahl nitrogen, total phosphorous (P), ammonium ($N\text{-}NH_4^+$), phosphate ($P\text{-}PO_4^{3-}$), and VFA.

The feedstock that arrived at the WWTP had already been mixed with the acidogenic fermentative inoculum, which initiated solubilizing and converting the organic solid matters into SCOD and VFAs in the transporter. Then, in the pretreatment unit, a sodium hydroxide (NaOH) solution (40% kg/kg) was added to bring the pH to 9 - 10 and heated to 60 ˚C for 24 hours. Subsequently, the biomixture was fed manually into a 5 L (operational volume of 4.5 L) continuously stirred pilot acidogenic fermenter operated at the given conditions (Table 1). Its high alkalinity maintained the pH during the acidogenic fermentation in the optimal range. Further, the mixture was blended mechanically, and the whole system was kept in the oven to hold the temperature constant at 37 ˚C. The output was sampled frequently during the week, and the samples were centrifuged to obtain the supernatant to measure pH, SCOD, VFA, $N\text{-}NH_4^+$, and $P\text{-}PO_4^{3-}$. A tiny fraction of the residual solid part was used to characterize solids like COD, P, and total Kjeldahl nitrogen, and the rest was kept in the freezer to apply the bio-methane potential (BMP) test.

Table . The operational parameters of the mesophilic acidogenic fermenter.

Hydraulic retention time (days)	Organic loading rate (kg–volatile solids/m^3.d)	Temperature (˚C)	pH[a], mean (SD)
4.5	6.89	37	6.56 (0.25)
3	10.33	37	6.7 (0.45)

[a]13 measurements for pH.

The VS and TS characterization were performed in 105 ˚C and 550 ˚C ovens for 24 hours, respectively. Except for VFAs, all the remaining analyses (including COD measurements) followed the standard methods for examining water and wastewater [25]. The methods described in the A and D sections of No. 5220 for COD quantification were used. These methods are named "Closed Reflux, Titrimetric Method" and "Closed Reflux, Colorimetric Method" for the solid and liquid phases, respectively [25]. For the liquid, the samples were filtered after being centrifuged at 4500 rounds per minute (rpm) for 5 minutes, and before the analysis, the supernatant was filtered with a 0.45 µm cellulose filter (Whatman). For the solid, acidic digestion was performed at 220 ˚C with a high pressure of 2 atmospheres to destroy the 0.2 g of solid matrix for 2 hours. Afterward, the COD was measured in the solution using titration by ferrous ammonium sulfate as described in the standard methods. Our limit of detection was 50 - 500 mg-COD/L for the calorimetric method and 40 - 400 mg-COD/L for the titrimetric method. In this research, dilution was done for high-concentration values that are beyond the considered limit of detection.

In the BMP test, the effect of biochar addition was observed for 3 diverse dosages (0, 0.12, and 0.24 g-biochar/g-VS) on the bio-methane volume, content, and production kinetics in the mesophilic condition using four sets of the BMP test. The tests were conducted with a total number of 8 bottles of 250 mL (working volume of 215 mL). The anaerobic condition was ensured in bottles by sealing them after filling without any flushing with nitrogen molecules (N_2) or carbon dioxide (CO_2) since we had known that oxygen transfer at the surface of the waste stream was impossible as it contained a high TS and SCOD. This type of procedure was adopted in our laboratory and has been conducted for years. The biochar was synthesized by a local supplier, and its main physical and chemical features are reported in Table S1 in Multimedia Appendix 1. It was ground into microparticles and kept under a dried condition at room temperature before being added to the bottles. Further, the inoculum for the BMP test was collected from the 2300 m^3 completely stirred anaerobic digester treating thickened WS and squeezed OMSW mixture under the mesophilic condition at an organic loading rate (OLR) of 1.8 - 2.0 kg-VS/m^3.d in the treatment plant. The inoculum was added to the feedstock (residual solid from acidogenic fermentation) based on the weight ratio of 0.3 FS g-VS/IN g-VS. The TS and VS contents in the bottles (ie, inoculum and feedstock) were 133 g/kg and 17.6 g/kg, respectively.

The experiments were conducted for each condition, namely, only inoculum and either with or without biochar, in 2 bottles. The test was terminated after 25 days when the cumulative biogas production reached almost 89% of the final projected value. The biogas content was characterized by gas chromatography (for days 1, 4, 6, 10, 14, 16, 18, 21, and 25). Additionally, the values for the remaining days were filled through imputation using the *k*-nearest neighbors algorithm (number of neighbors=4 and weights=distance) [26]. The imputation code is provided in the repository [27]. Then, the biogas and bio-methane volumes were subtracted from the only inoculum to correct for the endogenous methane production, and both values were averaged for 2 bottles. Gas chromatography was performed using Agilent Technology (TM 6890N) with an HP-PLOT MoleSieve column (30 m length, 0.53 mm ID × 25 mm film thickness) and a thermal conductivity detector with argon as a carrier (79 mL/min). The hydrogen molecule (H_2), CH_4, oxygen molecule (O_2), and N_2 were analyzed using a thermal conductivity detector at 250 ˚C. The inlet temperature was 120 ˚C, with constant pressure in the injection port (ie, 70 kilopascal [kPa]). Samples were taken using a gas-type syringe (200 µL). Once the entire sample was vaporized, peak separation occurred within the column at a constant temperature of 40 ˚C for 8 minutes. We did not plan to monitor pH and other parameters like alkalinity, VFA, ammonia, and phosphate because the pH drop risk was negligible, and the biochar addition could provide a buffer capacity and adsorption of inhibitory compounds in the solution [28]. Moreover, a considerable part of the readily biodegradable COD of the feedstock was already converted to VFAs in the previous step. As a result, the process was easily controlled even in the transient condition when the risk of methanogenic inhibition was high [29].

Statistical Analysis and Performance Indicators

The performance indicators, including COD solubilization, VFA yield, ammonia and phosphate release, and VFA/SCOD ratio were determined. These indicators characterize the mesophilic acidogenic fermentation on the days when the data were available, and the process reached the pseudo-steady state condition. The indicators were calculated, and the data were plotted using a Microsoft Excel spreadsheet (Version 2412). In addition, the VFA weight ratio distribution was determined from the total VFA weight on the same day. The process stability was evaluated based on variations in daily VFA concentrations. The formula for the performance parameters is reported in Multimedia Appendix 1. The exploratory data analysis and 2-tailed *t* test on VFA data were performed for the VFA concentration, yield, and VFA/SCOD ratio for 2 HRTs by the open source program R (version 3.5.0; The R Foundation for Statistical Computing). We assumed that the 2 datasets were paired and had a normal distribution. The code is provided in the repository [27]. The values for the 2 HRTs to increase the VFA concentration in the outlet were selected based on our experience and process knowledge. According to this information, exceeding the HRT value by more than 3 - 5 days can bring the process into an anaerobic digestion step. As a result, the VFAs with high added-value markets are converted to biogas. Hence, the 2 HRTs of 3 days and 4.5 days were tried in the pilot test, knowing that the VFA concentration would either increase or decrease linearly in this local region of operation.

For the BMP tests, two kinetic models were calibrated, namely, the first-order rate and modified Gompertz, to the biogases' cumulative yield. Additionally, the specific methane production (SMP) and specific biogas production (SGP) plus maximum volumetric methane content (v/v %) were determined. Comparing these results could reveal how the biochar addition, FS/IN ratio of 0.3, and pretreatment improved the process in terms of the rate and fostered methanogenesis. Such improvements are manifested through a higher hydrolysis rate,

a shorter lag phase, and a higher maximum volumetric methane content. Besides, the biogas yield was determined as g-biogas/g-VS.

Technical and Economic Assessment

This research sets up a mass flow analysis with parameters adopted for a municipality with 70,000 population equivalents (PEs) for the two scenarios: (1) a line with pretreatment and mesophilic acidogenic fermentation followed by mesophilic anaerobic digestion and (2) a single-stage mesophilic anaerobic digestion as currently deployed at the Treviso WWTP. This study focuses on water and energy preservation and increased profits from VFA production in our conversion line through several refinements. They were tied with the HRT identified in the previous step, integration of our process knowledge of using the fine-tuned FS/IN ratio, and biochar addition in anaerobic digestion. Detailed information and calculations regarding the mass flow analysis are available in the supplementary documents in the Excel spreadsheet named “Mass Balance” [27]. The following paragraph provides the full description of the two scenarios.

The two scenarios shared the first part of the model where the separated OMSW by a door-to-door collection system that was screw-pressed and diluted with water to reach the TS of 280 g/kg. Then, in the first scenario, adding a sodium hydroxide solution (40% kg/kg) elevated the feedstock pH to 9 - 10. Afterward, the solution was heated at 60 ˚C for 24 hours in the pretreatment unit. Next, it was diluted and heated further before feeding into the mesophilic acidogenic fermenter based on the desirable HRT. The last part of the first scenario was the optimized anaerobic digestion of residual fermented solids. Specifically, the stability endowment by adding biochar to the anaerobic digestion could ultimately smooth running the process in a high OLR (low water dilution). Furthermore, an FS/IN ratio of 0.3 was applied to increase the kinetics rate with the benefit of a decrease in digester volume, energy consumption, and capital cost. This finding is of significant importance in plants and zones with limited area, water, and energy.

In the second scenario, the screw-pressed feedstock was diluted and immediately fed into a mesophilic anaerobic digester for only biogas production.

It was assumed that the reactors transfer heat from the walls with the atmosphere and earth. Further, the biogas would be consumed in the combined heat and power units for electricity production with an overall efficiency of 0.4. In this research, the mass of VFAs and the net amount of energy production were accounted for as the source of income. Meanwhile, the corresponding costs were the operational expenditure, the mass of the water process, and the final residual solids to dispose of. Reference parameters for the energy analysis and boundary conditions are given in Table 2. The price of electricity was assumed to be €130 per megawatt-hour (MWh). These two scenarios were compared to identify the most favorable one regarding surplus thermal energy and electricity or the shorter payback period.

Table . Reference parameters and boundary conditions for energy flow analysis.

Parameter	Heat transfer coefficient ($W/(m^2.^\circ C)$)	Temperature (˚C)	Low heat value (MJ^a/Nm^3)	Energy conversion efficiency
Biogas	—[b]	—	23.012	—
Thermal energy yield	—	—	—	0.5
Electrical energy yield	—	—	—	0.4
Operative temperature	—	37	—	—
Water temperature	—	15	—	—
Air temperature	—	20	—	—
Ground temperature	—	25	—	—
Outer concrete reactor wall	0.7	—	—	—
Inner concrete reactor wall	1.2	—	—	—
Floor	2.85	—	—	—

[a]MJ: megajoules.

[b]Not applicable.

Ethical Considerations

This research was not conducted on human or animal subjects and does not involve the collection of any new data. Therefore, it was unnecessary to obtain ethics approval.

Results

Biorefinery Process Scheme and Experimental Studies: Composition and Characteristics of the Pretreated Feedstock

The pretreated feedstock’s main physical and chemical characteristics were quite stable throughout the experiment (Table 3). The feedstock had an average TS content of 45 (SD 3.15) g/kg and VS content of 32 (SD 3.28) g/kg. These values

suggest that the biodegradable solids constituted 72% of the TS, which could support the fermentation process. The chemical composition of the solid part was 12.9 g-N/kg-TS, 4 g-P/kg-TS, and 565 g-COD/kg-TS, which was in the range of the values reported for the typical OMSW in Italy [30]. The chemical composition of the liquid was 325 mg N-NH_4^+/L, 14 mg P-PO_4^{3-}/L, and 25.8 g-SCOD/L. Further, the feedstock COD:N:P ratio was determined as 100:2.2:0.7, meaning that nutrients such as phosphor and nitrogen should not be the limiting substrates in acidogenic fermentation [31]. In this regard, the slight level of VFA concentration at the level of 3.5 g-SCOD/L was due to acidogenic fermentation, which had been happening during transportation.

Table . Main physical-chemical features of the feedstock.

Parameter	Weight ratio (g/kg)	Mass ratio (%)	Concentration (mg/L)
Total solids, mean (SD)[a]	45 (3.15)	—[b]	—
Volatile solids, mean (SD)[a]	32 (3.28)	—	—
Total Kjeldahl nitrogen[c]	12.9	—	—
Phosphorous[c]	4	—	—
Chemical oxygen demand[c]	565	—	—
Chemical oxygen demand:nitrogen:phosphorous	—	100:2.2:0.7	—
Soluble chemical oxygen demand	—	—	25,814
N-NH_4^+[d]	—	—	325
P-PO_4^{3-}[e]	—	—	14
Volatile fatty acid[c]	—	—	3500
Volatile solids/total solids, mean (SD)[a]	—	72 (5)	—

[a]Based on 3 measurements.

[b]Not applicable.

[c]Measurements done for nitrogen, phosphor, and soluble chemical oxygen demand equivalents for total Kjeldahl nitrogen, phosphorous, chemical oxygen demand, and volatile fatty acid.

[d]N-NH_4^+: ammonium.

[e]P-PO_4^{3-}: phosphate.

Statistical Analysis and Performance Indicators

Acidogenic Fermentation

Table 4 presents the main physical and chemical characteristics of the effluent and solid cake from the acidogenic fermenters. According to Figure 2, the process reached a steady condition after 14 days, which was roughly 3 times the HRT (4.5 days). Both HRTs were similarly stable in terms of VFA concentration variation because of a negligible difference between SDs: 2.82 g-SCOD/L versus 2.45 g-SCOD/L. These values are less than 10% of the total VFA, and the VFA production continued for more than 3 weeks without any considerable issues. The lack of any change in this process is attributed to the initial high pH of 9 - 10, which supported the process by keeping the pH variation in the optimal range of 6 - 7.5 [32].

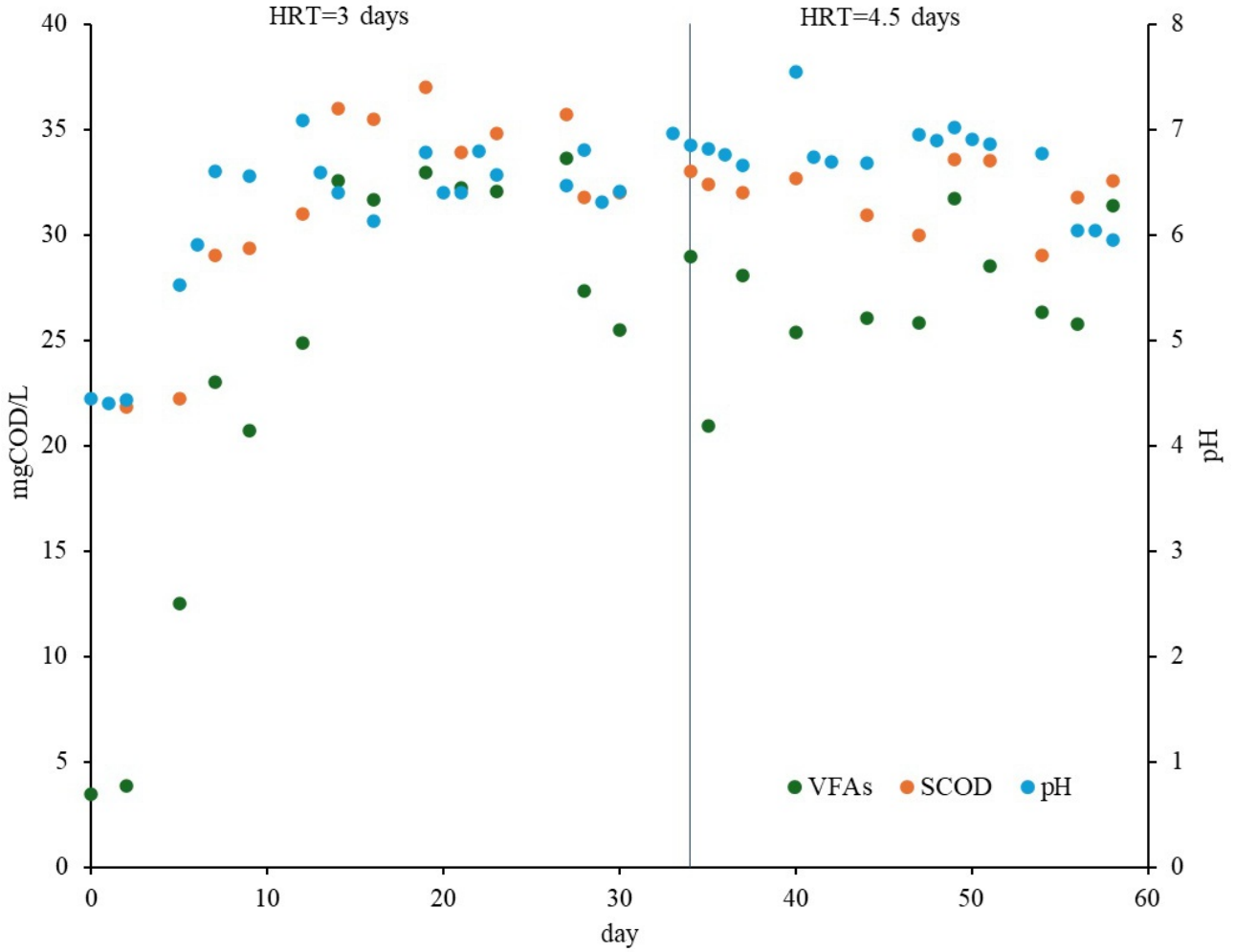
Table . Main physical-chemical features of the effluent and solid cake from mesophilic acidogenic fermentation.

Hydraulic retention time (days)	Total solids (g/kg), mean (SD)	Volatile solids (g/kg), mean (SD)	Volatile fatty acid (g–soluble chemical oxygen demand/L), mean (SD)	pH, mean (SD)
4.5[a]	43 (5.15)	23.6 (2.07)	30.77 (2.82)	6.56 (0.25)
3[b]	38 (4.55)	25.8 (1.5)	27.67 (2.45)	6.7 (0.45)

[a]5 measurements for total solids and volatile solids; 9 measurements for volatile fatty acid; 13 measurements for pH.

[b]4 measurements for total solids and volatile solids; 9 measurements for volatile fatty acid; 13 measurements for pH.

Figure 2. VFA, SCOD, and pH for mesophilic acidogenic fermentation. COD: chemical oxygen demand; HRT: hydraulic retention time; SCOD: soluble chemical oxygen demand; VFA: volatile fatty acid.



Based on the *t* test results (t_8=−2.68; *P*=.03; CI=95%), it was verified that the mean VFA concentration for an HRT of 4.5 days was significantly higher than the value for 3 days (30.77 vs 27.67 g-SCOD/L). A similar statistical analysis (t_8=−0.99; *P*=.35; CI=95%) for the VFA/SCOD ratio rejected the significance of a higher mean value of 0.892 (SD 0.04) for an HRT of 4.5 days than 3 days, with a mean value of 0.87 (SD 0.058). The possible range of values for the VFA concentrations and VFA/SCOD, which cover 99% and 50% of the data for the 2 HRTs, are depicted by the box plots in Figures 3 and 4, respectively.

Figure 3. Box plot of volatile fatty acid concentrations for mesophilic acidogenic fermentation. HRT: hydraulic retention time; SCOD: soluble chemical oxygen demand.

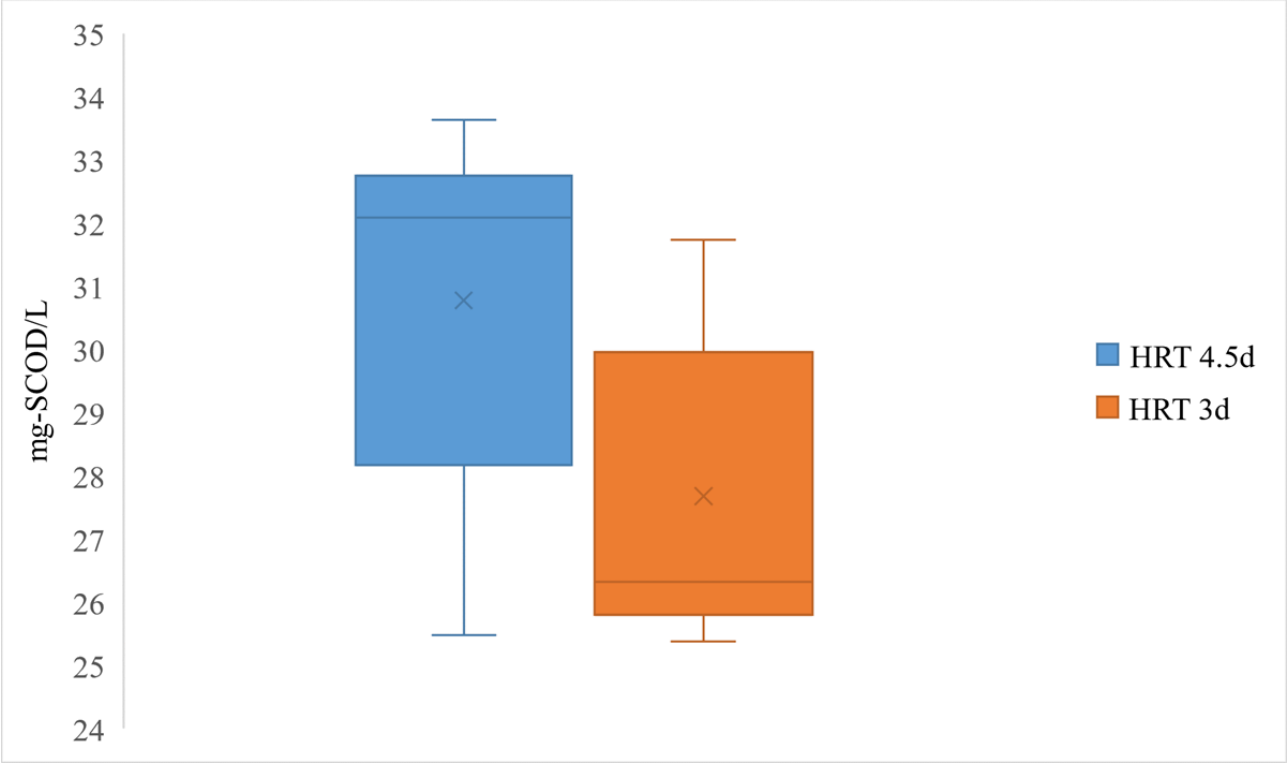
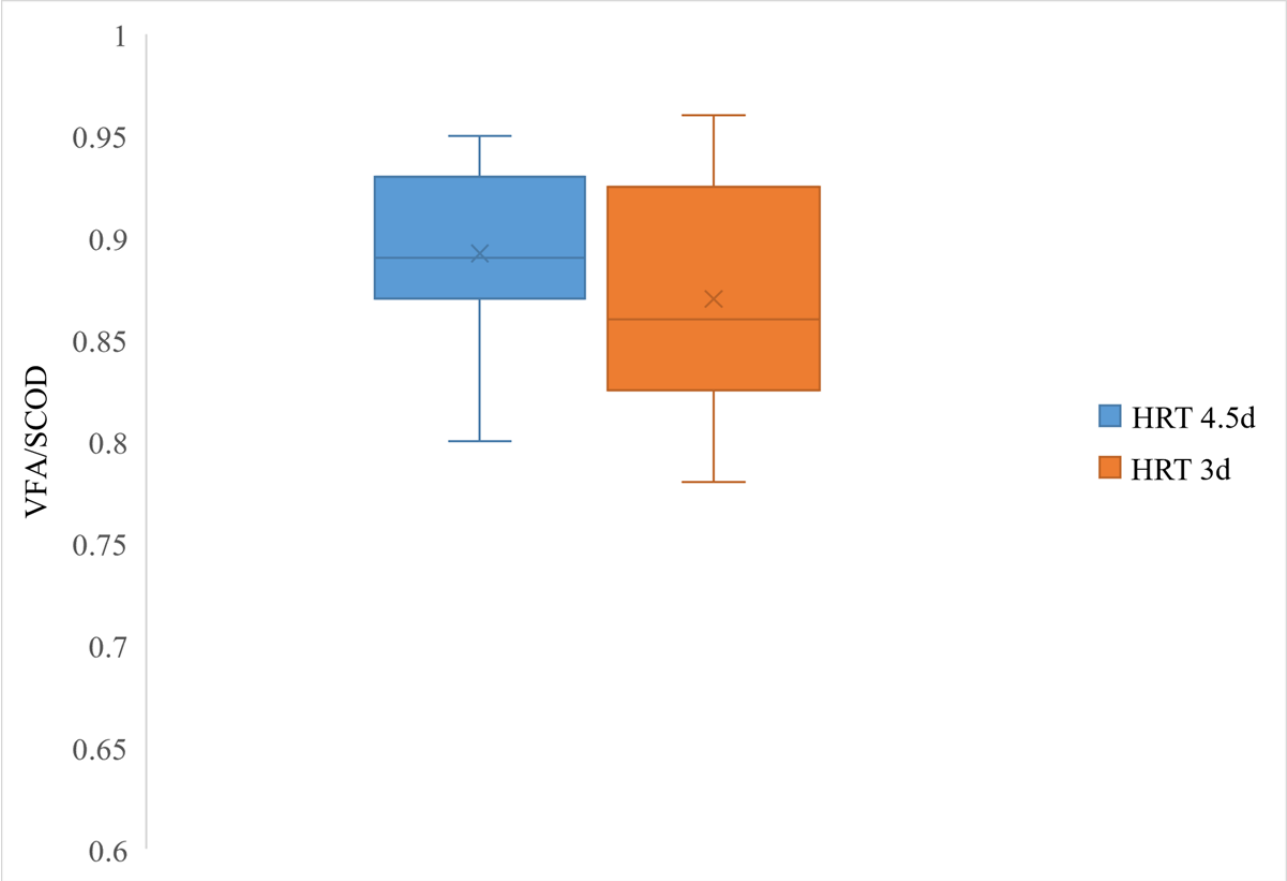


Figure 4. Box plot of VFA/SCOD ratios for mesophilic acidogenic fermentation. HRT: hydraulic retention time; SCOD: soluble chemical oxygen demand; VFA: volatile fatty acid.



Performance parameters for the 2 HRTs are given in Table 5. As can be seen, the HRT of 4.5 days gave higher COD solubilization and released more ammonia and phosphate than the HRT of 3 days. Moreover, the 0.57 VFA yield per gram of

VS for the HRT of 4.5 days was significantly higher than 0.5 for the HRT of 3 days (t_8=−2.94; P=.02; CI=95%).

In the biopolymer-synthesizing process, the aim was to generate a stable VFA weight ratio distribution with a high VFA/SCOD ratio for an efficient PHA synthesis during the whole process. Concisely, the VFA stream with a higher dominance of even numbers of carbon atom acids means a higher 3-hydroxybutyrate monomer synthesis compared to the 3-hydroxyvalerate, which is correlated with the net prevalence of odd numbers of carbon atom acids (propionic, valeric, and isovaleric acid) [33]. As can be inferred, the stability in the VFA spectrum means a predictable and reproducible PHA monomer production. Accordingly, the physical and mechanical features of synthesized biopolymers are stable [34,35].

Figure 5 reports the weight ratio distribution of the VFAs for the 2 HRTs. The main fractions were acetic acid (38% - 42%), butyric acid (24%), caproic acid (16% - 18.5%), propionic acid (9% - 11%), and valeric acid (5%). This VFA distribution, with a major part of butyric and acetic acid, is in line with those reported in similar studies [29,31]. In this respect, the VFA weight ratio distribution is determined by the type of feedstock and food waste rather than the operational conditions.

Table . Performance parameters of two different operational conditions used in mesophilic acidogenic fermentation.

Hydraulic retention time (days)	Solubilization (Δg–soluble chemical oxygen demand/g-VS^a_0), mean (SD)	Y_{VFA}[b] (Δg-VFA/g-VS_0), mean (SD)	Ammonia release (%), mean (SD)	Phosphate release (%), mean (SD)
4.5[c]	0.28 (0.06)	0.57 (0.06)	35 (10.74)	13.7 (8.77)
3[d]	0.19 (0.05)	0.50 (0.06)	29 (0.11)	11 (0.06)

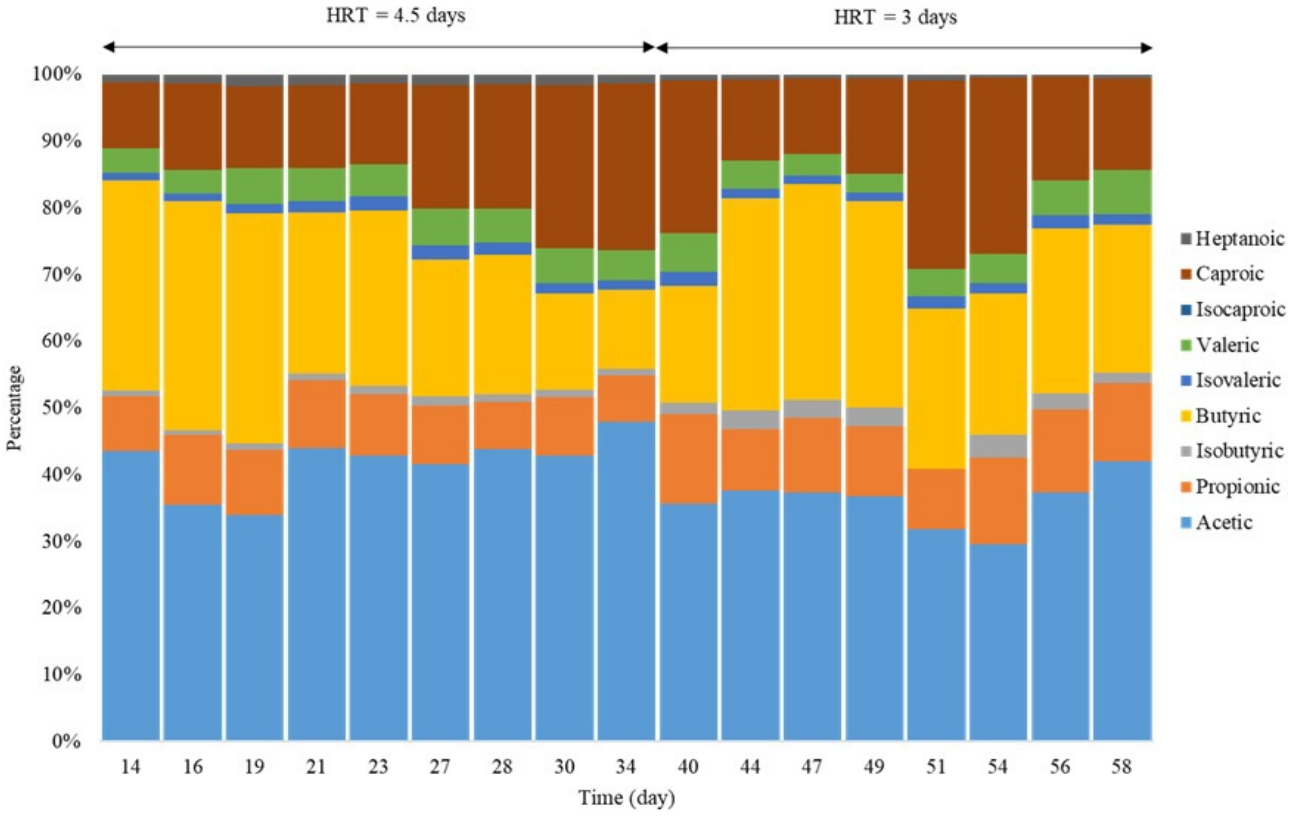
[a]VS: volatile solids.

[b]VFA: volatile fatty acid.

[c]9 measurements for solubilization (Y_{VFA}); 8 measurements for ammonia and phosphate release.

[d]9 measurements for solubilization (Y_{VFA}); 7 measurements for ammonia and phosphate release.

Figure 5. Volatile fatty acid weight ratio distribution for mesophilic acidogenic fermentation. HRT: hydraulic retention time.



Anaerobic Digestion

Table 6 summarizes the performance parameters and the results from the kinetics study for anaerobic digestion. This study obtained a remarkably high value for the hydrolysis rate (ie, 0.58, 1/d) with no lag phase. Besides, a biogas yield of 0.61 - 0.83 g-biogas/g-VS, SMP of 0.133 - 0.204 CH_4-Nm^3/kg-VS, and an average composition of 45% - 58% methane (v-CH_4/v-biogas) were obtained. According to Figure 6, adding biochar provided the desirable conditions for the growth of hydrogen using methanogenesis manifested through

a higher maximum volumetric methane content (86% vs 66% volumetric basis [v/v]).

Table . The performance indicators for anaerobic digestion and results from the kinetics study for two models: (1) first-order rate and (2) modified Gompertz.

Experiments	Specific methane production (CH_4[a]-Nm^3/kg-VS[b])	Specific gas production (CH_4-Nm^3/kg-VS)	K[c] (1/d)	R_m[d] (CH_4-mL/g-VS.d)	[e] (days)	RMSE[f] first-order (CH_4-Nm^3/kg-VS)	RMSE modified Gompertz (CH_4-Nm^3/kg-VS)	Max CH_4 content (v/v[g]), %
Without biochar	0.204	0.540	0.57	76.12	0	10.4	6.82	68.5
Biochar (0.12 g-biochar/g-)	0.133	0.567	0.69	62.42	0	5.74	5.59	86
Biochar (0.24 g-biochar/g-)	0.177	0.500	0.58	65.17	0	9.64	3.39	76.5

[a]CH_4: methane.

[b]VS: volatile solids.

[c]Hydrolysis rate.

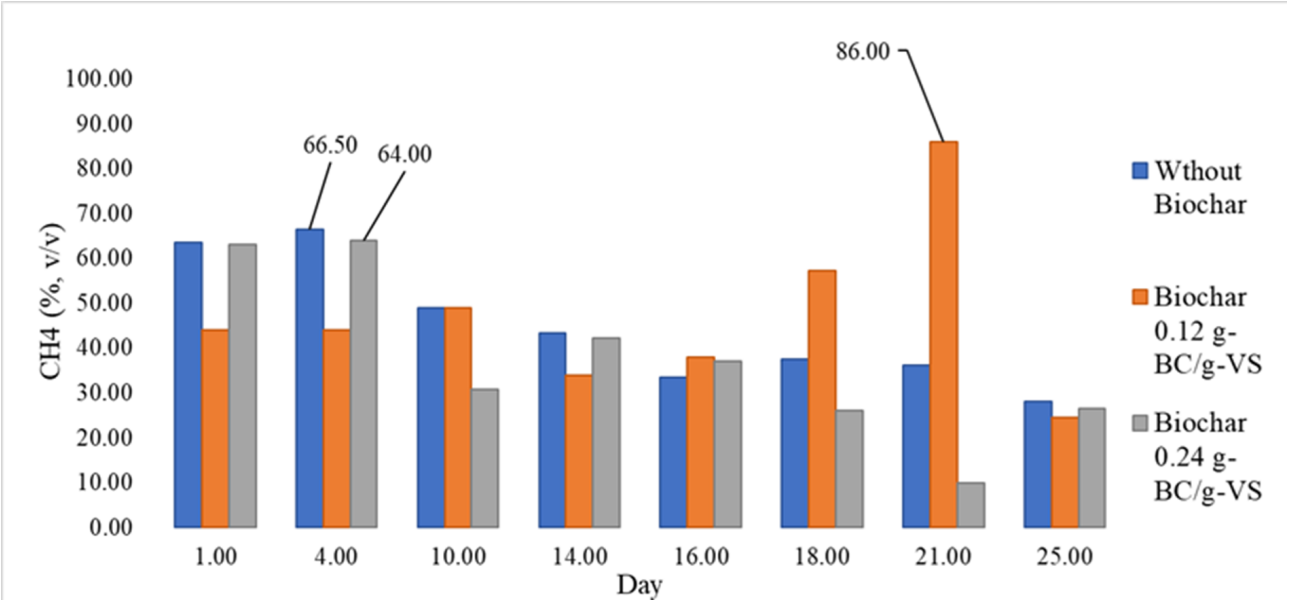
[d]Maximum methane production rate.

[e]Lag phase.

[f]RMSE: root mean squared error.

[g]v/v: volumetric basis.

Figure 6. CH_4 content in v/v for 3 different biochar dosages in anaerobic digestion. BC: biochar; CH_4: methane; VS: volatile solids; v/v: volumetric basis.



The mass flow model was adopted for 0.12 g-biochar/g-VS as the only feasible solution. Unlike other dosages, it could satisfy the assumptions for an FS/IN ratio of 0.3 at an HRT of 20 days, which was adequately long enough to let the methanogens reproduce themselves. Detailed information is available in the Excel sheet named "DIGESTER DESIGN" [27]. Besides, the high alkalinity of the biochar as reported in Table S1 in Multimedia Appendix 1 signifies a benefit of the biochar addition in limiting the concern about decreases in pH for a high OLR in full-scale implementation. Accordingly, almost 4-fold of the ordinary OLR was obtained, that is, 6.25 kg-VS/m^3.d, by minimum water dilution, knowing that the biochar could maintain the stability of the process. Therefore, the digester volume will decline at the rate of 28 L/PE. Hence, the presented mass flow line model was implemented based on the results of 0.12 g-biochar/g-VS, the weighted average composition of biomethane as 35% v/v, and the SGP as 0.56 biogas-Nm^3/kg-VS for an HRT of 20 days corresponding to an FS/IN ratio of 0.3.

Based on the root mean squared error reported in Table 6, both models were almost identical in describing biomethane production for a biochar dosage of 0.12 g-biochar/g-VS, and for simplicity, we used the first-order rate model in the feasibility study.

Technical and Economic Assessment

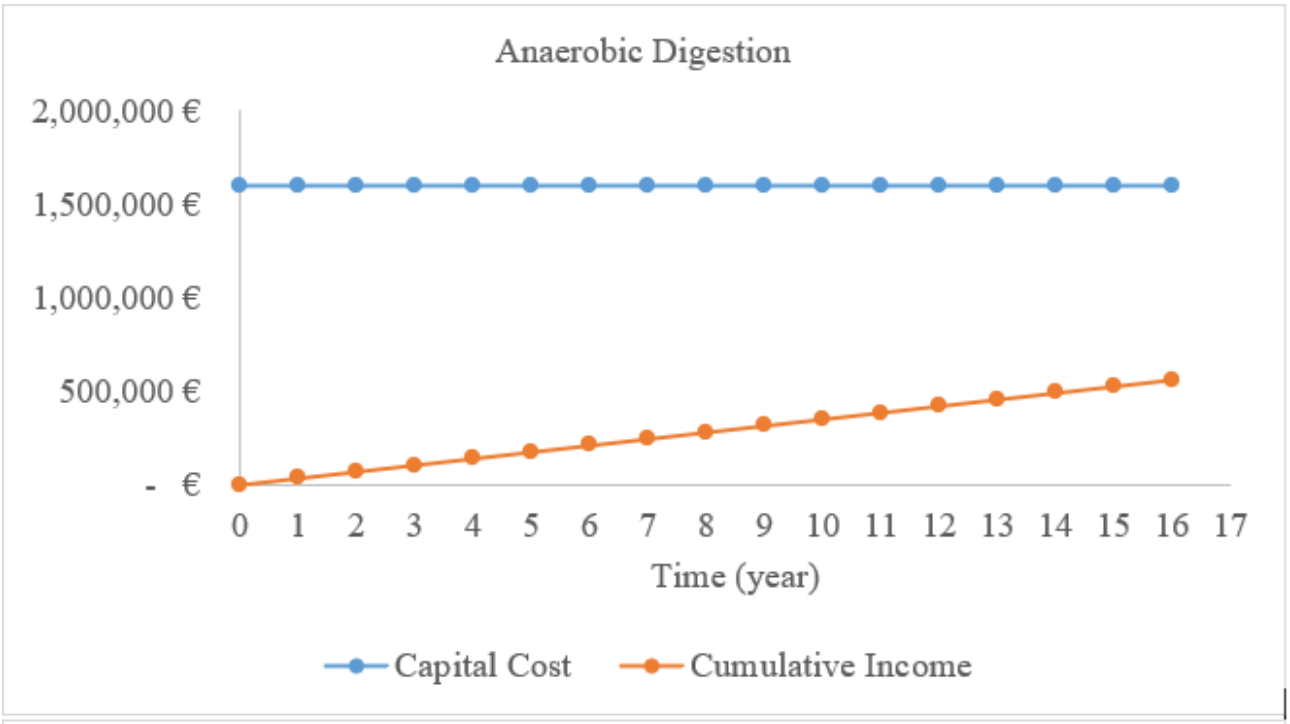
Assuming an imaginary municipality of 70,000 PEs and the amount of TS production per capita as 0.3 kg/PE per day [36], the inlet to the scale-up line would be 21,000 kg-TS per day.

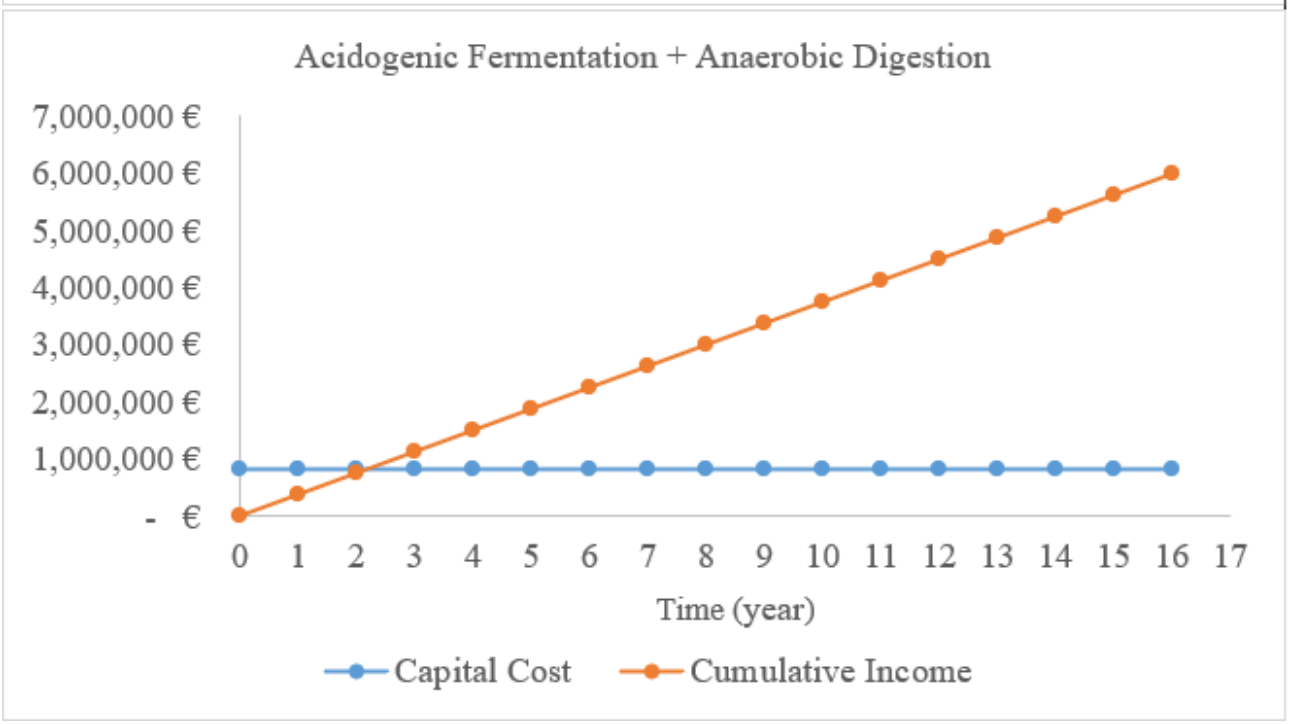
In the first scenario, the biowaste stream, after passing through the screw press and pretreatment unit, had a mass flow of 113,788 kg per day, TS of 4.1% kg/kg, and VS of 3.1% kg/kg. Then, the mixture was heated to 37 ˚C before and in the acidogenic fermenter, which was operated at an HRT of 4.5 days and OLR of 6.89 kg-VS/m^3.d. This process was performed to convert biosolids into the VFAs and SCOD at concentration levels of 30.77 g-SCOD/L and 34 g-SCOD/L, respectively. At this step, the gaseous flow rate was assumed to be zero, as an HRT of 4.5 days is short for any adequate growth of methanogens in mesophilic conditions. The stream out of the acidogenic fermenter had a mass flow rate of 113,788 kg per day, with a VFA content of 3501 kg-SCOD per day, which could be used in the PHA-synthesizing step [37]. The outlet of this step was used in the separator to gain overflow and solid cake. Later, the solid cake was minimally diluted by water before being fed into a mesophilic anaerobic digester with a biochar addition of 0.12 g-biochar/g-VS. The anaerobic digester received a TS content of 18% kg/kg and a flow rate of 18,180 kg per day, corresponding to an HRT of 20 days and OLR of 6.25 kg-VS/m^3.d. Overall, an SGP of 0.285 (Nm^3-biogas/kg-VS) was obtained assuming zero gas production in acidogenic fermentation.

In the second scenario, the fresh feedstock, after being screw-pressed, had a mass flow rate of 4678 kg-TS per day and 28% kg/kg dry matter. Then, it was diluted with water and heated before being fed into the anaerobic digester. At this step, the mass flow rate of 85,012 kg per day with a TS of 6% kg/kg entered the digester with a volume of 2125 m^3, leading to an HRT of 25 days and OLR of 1.7 kg-VS/m^3.d. The SMP of 0.311 Nm^3-biogas/kg-VS was obtained by destroying 80% of the VS.

In this study, working volumes of 512 m^3 and 364 m^3 were adopted for the acidogenic fermenter and anaerobic digester in the first scenario, respectively, and 2125 m^3 for the anaerobic digester in the second scenario. As a result, the capital cost for the presented line was almost €809,000, roughly half of the quantity for the single-step anaerobic digestion (Figure 7). Unlike the single-step anaerobic digestion that converts all VS to biogas, this novel line shared the recovery of VS between higher added-value VFAs and biogas production, and expectedly generated 10-fold higher benefits (€375,085). Consequently, the payback period was reduced by more than 20 times in 2 years (Figure 7). This period was achieved using less surplus energy (2251 megajoules [MJ]/d) for the 2-step fermentation (vs 21,567 MJ/d for the single-step anaerobic digestion).

Figure 7. Capital cost and cumulative yearly income for the two proposed scenarios. A currency exchange rate of €1=US $1.05 is applicable.





Discussion

Principal Results

We showed that multistep fermentation followed by anaerobic digestion is both economically and technically feasible. The findings indicated that producing VFAs and biogas in separate stages can significantly reduce the payback period for upcoming investments in biorefinery projects and result in the creation of a highly desired stream that is rich in VFAs. Additionally, the process stability could be maintained even at a high OLR by adding biochar and converting the VS's easily biodegradable COD content into VFAs in the first phase. This would preserve energy and water, and reduce the digester's volume.

Comparison With Previous Studies

Because of the extra pretreatment unit in this research, the VFA yield of 0.57 - 0.63 Δg-VFA/g-COD_{IN} was roughly double the value reported by Valentino et al [31] for the same OMSW.

Our results also indicate a substantial improvement in the process kinetics, which was manifested through a more than 8-fold rise in the hydrolysis rate (0.58 vs 0.07, 1/d) and a full decrease in the lag phase (0 vs 2.69 days) as opposed to the previous study by Karki et al [38]. This improvement is

attributed to the destruction of the solids structure caused by bacterial enzymes and a hot alkaline solution. Additionally, a higher active biomass per feedstock was provided using a fine-tuned FS/IN ratio of 0.3 (VS basis), which was noticeably lower than the quantities (1 and 0.5) reported in similar studies [38,39].

The values for SMP and mean methane volumetric content presented in this study are lower than those reported by Valentino et al [29] (ie, 0.25 CH_4-Nm^3/kg-VS and 63% - 64% v/v, respectively). This difference is explained by the added fresh WS, which has a higher digestible content and better nutrient balance than the fermented solids. Similarly, the SMP in this study was lower than the 0.384 CH_4-Nm^3/kg-VS found in the study by Moreno et al [39]. This study investigated the anaerobic digestion of residual solids from two steps of bioethanol production and saccharification on OMSW. In this respect, bioethanol production can only convert part of the cellulose, starch, and some dissolved carbohydrates. Consequently, a great part of the biosolids' volatile content, nearly 70%, is still available to be harvested in different biorefinery schemes compared with the one proposed in this method with 55%. Besides, the fermented OMSW would have a completely incompatible composition since it did not only come from different geographical locations (Spain and the United Kingdom) with different food habits but also underwent different biological pretreatment. Further, the multistep recovery line proposed in our study is more practicable technically. As the method studied by Moreno et al [39] requires sterilization conditions, imposing an additional operational cost and bioethanol concentration should be high enough to lower the cost of the subsequent distillation step.

Furthermore, our method for VFA production distinctively from biogas was preferable to the study by Papa et al [9], wherein the operational alteration on a single anaerobic digester was performed to obtain VFAs and biogas. These researchers asserted that the single-step recovery of biogas and VFAs was feasible by increasing the OLR while keeping the SMP of the reactor almost unaffected. The main recovery path for the VS was still biogas production in their study, which accounted for more than 90% of the VS conversion. Meanwhile, our study obtained 36% and 64% of the biogas and VFA conversion share, respectively. Further, whereas the destruction of VS of around 70% was achieved in both studies, their proposal limited the VFA distribution to propionic and butyric acid. The explanation is that some of the VFAs were converted into biogas in the same unit, which could negatively affect the PHA synthesis step later.

Conclusion and Limitations

This paper demonstrated the technical and economic feasibility of a multistep recovery line for OMSW. The results of this study indicate that the production of VFAs and biogas in distinct steps can considerably shorten the payback period for future investments in biorefinery projects and produce a highly desirable VFA-rich stream. Further, adding biochar and converting easily degradable COD content in the VS into VFAs in the first step could maintain the process stability even with a high OLR in anaerobic digestion. As a result, it leads to energy and water preservation and a decrease in the digester volume. However, consideration should be paid to the full-scale implementation since the pilot studies cannot resemble the stability of the real process. For instance, operational alterations such as raising the OLR and the addition of biochar in the full-scale implementation might perturb the process pH or the synergetic balance between the bacterial communities and stop the process completely, which was never observed in our experimental study. Further, the superb profitability of the proposed line was highly variable because our cost analysis was too simplistic and did not elaborate on all the possible associated expenditures and incomes. Besides, since many of its components were from subject matter experts rather than the pilot studies' budget, they were prone to site variations and uncertainties. Addressing the systematic uncertainty in the labor and material costs due to the changes in the supply chain issues, inflation, and site variations is beyond our scope. Moreover, caution should also be considered regarding the significance of the BMP results with the marginal difference since the number of samples was not large enough for statistical analysis. Nevertheless, the results presented in this study were prepared cautiously both technically and financially to encourage the revolution in the current state of organic waste valorization in Italy and any similar location.

In conclusion, a robust framework was proposed to assess the valorization of organic waste through experimental tests, statistical analysis, process kinetics, and mass and energy flow analysis. The findings support considerably higher profitability and, thus, a shorter payback period for the multistep fermentation than the current single anaerobic digestion. Additionally, our results encourage the circular economy perspective on converting OMSW into biogas and VFAs with the benefit of fewer residual solids due to reusing them in a pyrolysis line.

Acknowledgments

The author gratefully acknowledges the Italian Ministry of University and Research and the University of Ca′ Foscari for financially supporting and providing a study design for this research in the frame of Programma Operativo Nazionale Ricerca e Innovazione 2014-2020. In addition, the author appreciates Marco Gottardo for his helpful assistance with gas chromatography, and Francesco Valentino for his helpful comments and guidance during the study, as well as Alessio Dell ′Olivo and Aditi Parmar Chitharanjan for their helpful advice with laboratory experiments and material flow analysis. The author also attests that they used generative artificial intelligence in checking the text regarding grammar and correct word usage.

Data Availability

The codes for statistical analysis as well as the datasets generated and analyzed during this study are available from a repository [27]. Research Data Policy Type 2 (for life sciences) by Springer Nature is distributed under the terms of the Creative Commons Attribution 4.0 International License.

Conflicts of Interest

The author declares his current expert witness position as a peer reviewer in the *Journal of Medical Internet Research*.

Multimedia Appendix 1
Supplementary table, equation, digester design, and code.
[DOCX File, 31 KB - xmed_v6i1e50458_app1.docx]

References

1. Preparatory study on food waste across EU 27: final report. Publications Office of the European Union. 2011. URL: https://data.europa.eu/doi/10.2779/85947 [accessed 2024-01-29]
2. Pfaltzgraff LA, De bruyn M, Cooper EC, Budarin V, Clark JH. Food waste biomass: a resource for high-value chemicals. Green Chem 2013;15(2):307-314. [doi: 10.1039/c2gc36978h]
3. Circular economy action plan. European Commission: Environment. 2022. URL: https://ec.europa.eu/environment/strategy/circular-economy-action-plan_en#:~:text [accessed 2025-01-23]
4. Mazur-Wierzbicka E. Towards circular economy—a comparative analysis of the countries of the European Union. Resources 2021;10(5):49. [doi: 10.3390/resources10050049]
5. Mattioli A, Gatti GB, Mattuzzi GP, Cecchi F, Bolzonella D. Co-digestion of the organic fraction of municipal solid waste and sludge improves the energy balance of wastewater treatment plants: Rovereto case study. Renewable Energy 2017 Dec;113:980-988. [doi: 10.1016/j.renene.2017.06.079]
6. Cabbai V, De Bortoli N, Goi D. Pilot plant experience on anaerobic codigestion of source selected OFMSW and sewage sludge. Waste Manag 2016 Mar;49:47-54. [doi: 10.1016/j.wasman.2015.12.014] [Medline: 26739455]
7. Girotto F, Alibardi L, Cossu R. Food waste generation and industrial uses: a review. Waste Manag 2015 Nov;45:32-41. [doi: 10.1016/j.wasman.2015.06.008] [Medline: 26130171]
8. Tamisa J, Lužkov K, Jiang Y, van Loosdrecht MCM, Kleerebezem R. Enrichment of Plasticicumulans acidivorans at pilot-scale for PHA production on industrial wastewater. J Biotechnol 2014 Dec 20;192 Pt A:161-169. [doi: 10.1016/j.jbiotec.2014.10.022] [Medline: 25456060]
9. Papa G, Pepè Sciarria T, Carrara A, Scaglia B, D'Imporzano G, Adani F. Implementing polyhydroxyalkanoates production to anaerobic digestion of organic fraction of municipal solid waste to diversify products and increase total energy recovery. Bioresour Technol 2020 Dec;318:124270. [doi: 10.1016/j.biortech.2020.124270] [Medline: 33099102]
10. Jung S, Shetti NP, Reddy KR, et al. Synthesis of different biofuels from livestock waste materials and their potential as sustainable feedstocks – a review. Energy Conversion Manage 2021 May;236:114038. [doi: 10.1016/j.enconman.2021.114038]
11. Pagliano G, Ventorino V, Panico A, Pepe O. Integrated systems for biopolymers and bioenergy production from organic waste and by-products: a review of microbial processes. Biotechnol Biofuels 2017 May 2;10:113. [doi: 10.1186/s13068-017-0802-4] [Medline: 28469708]
12. Sun J, Zhang L, Loh KC. Review and perspectives of enhanced volatile fatty acids production from acidogenic fermentation of lignocellulosic biomass wastes. Bioresour Bioprocess 2021 Aug 2;8(1):68. [doi: 10.1186/s40643-021-00420-3] [Medline: 38650255]
13. Sampath P, Brijesh, Reddy KR, et al. Biohydrogen production from organic waste – a review. Chem Eng Technol 2020 Jul;43(7):1240-1248. [doi: 10.1002/ceat.201900400]
14. Gottardo M, Dosta J, Cavinato C, et al. Boosting butyrate and hydrogen production in acidogenic fermentation of food waste and sewage sludge mixture: a pilot scale demonstration. J Cleaner Production 2023 Jun 10:136919. [doi: 10.1016/j.jclepro.2023.136919]
15. Micolucci F, Gottardo M, Bolzonella D, Pavan P, Majone M, Valentino F. Pilot-scale multi-purposes approach for volatile fatty acid production, hydrogen and methane from an automatic controlled anaerobic process. J Cleaner Production 2020 Dec 20;277:124297. [doi: 10.1016/j.jclepro.2020.124297]
16. Anderottola G, Canziani R, Foladori P, Ragazzi M, Tatano F. Laboratory scale experimentation for RBCOD production from OFMSW for BNR systems: results and kinetics. Environ Technol 2000 Dec;21(12):1413-1419. [doi: 10.1080/09593332208618173]
17. Stoyanova E, Lundaa T, Bochmann G, Fuchs W. Overcoming the bottlenecks of anaerobic digestion of olive mill solid waste by two-stage fermentation. Environ Technol 2017 Feb;38(4):394-405. [doi: 10.1080/09593330.2016.1196736] [Medline: 27279450]

18. Sauer M, Porro D, Mattanovich D, Branduardi P. Microbial production of organic acids: expanding the markets. Trends Biotechnol 2008 Feb;26(2):100-108. [doi: 10.1016/j.tibtech.2007.11.006] [Medline: 18191255]
19. Dahiya S, Sarkar O, Swamy YV, Venkata Mohan S. Acidogenic fermentation of food waste for volatile fatty acid production with co-generation of biohydrogen. Bioresour Technol 2015 Apr;182:103-113. [doi: 10.1016/j.biortech.2015.01.007] [Medline: 25682230]
20. Ma Y, Gu J, Liu Y. Evaluation of anaerobic digestion of food waste and waste activated sludge: soluble COD versus its chemical composition. Sci Total Environ 2018 Dec 1;643:21-27. [doi: 10.1016/j.scitotenv.2018.06.187] [Medline: 29935360]
21. Montalvo S, Vielma S, Borja R, Huiliñir C, Guerrero L. Increase in biogas production in anaerobic sludge digestion by combining aerobic hydrolysis and addition of metallic wastes. Renewable Energy 2018 Aug;123:541-548. [doi: 10.1016/j.renene.2018.02.004]
22. Kumar M, Xiong X, Sun Y, et al. Critical review on biochar‐supported catalysts for pollutant degradation and sustainable biorefinery. Adv Sustainable Syst 2020 Oct;4(10):1900149. [doi: 10.1002/adsu.201900149]
23. Kumar M, Dutta S, You S, et al. A critical review on biochar for enhancing biogas production from anaerobic digestion of food waste and sludge. J Clean Prod 2021 Oct 10;305:127143. [doi: 10.1016/j.jclepro.2021.127143] [Medline: 36570877]
24. Inyang M, Gao B, Pullammanappallil P, Ding W, Zimmerman AR. Biochar from anaerobically digested sugarcane bagasse. Bioresour Technol 2010 Nov;101(22):8868-8872. [doi: 10.1016/j.biortech.2010.06.088] [Medline: 20634061]
25. Lipps WC, Braun-Howland EB, Baxter TE, editors. Standard Methods for the Examination of Water and Wastewater, 24th edition: American Public Health Association; 2022.
26. Cinar S, Cinar S, Kuchta K. Machine learning algorithms for temperature management in the anaerobic digestion process. Fermentation 2022 Jan 30;8(2):65. [doi: 10.3390/fermentation8020065]
27. Borhany H. Supplementary documents for converting organic municipal solid waste into volatile fatty acids and biogas: experimental pilot and batch studies with statistical analysis. Zenodo. 2024 Jul 14. URL: https://zenodo.org/records/12739504 [accessed 2025-01-23]
28. Mumme J, Srocke F, Heeg K, Werner M. Use of biochars in anaerobic digestion. Bioresour Technol 2014 Jul;164:189-197. [doi: 10.1016/j.biortech.2014.05.008] [Medline: 24859210]
29. Valentino F, Moretto G, Gottardo M, Pavan P, Bolzonella D, Majone M. Novel routes for urban bio-waste management: a combined acidic fermentation and anaerobic digestion process for platform chemicals and biogas production. J Cleaner Production 2019 May 20;220:368-375. [doi: 10.1016/j.jclepro.2019.02.102]
30. Moretto G, Valentino F, Pavan P, Majone M, Bolzonella D. Optimization of urban waste fermentation for volatile fatty acids production. Waste Manag 2019 Jun 1;92:21-29. [doi: 10.1016/j.wasman.2019.05.010] [Medline: 31160023]
31. Valentino F, Gottardo M, Micolucci F, et al. Organic fraction of municipal solid waste recovery by conversion into added-value polyhydroxyalkanoates and biogas. ACS Sustainable Chem Eng 2018 Dec 3;6(12):16375-16385. [doi: 10.1021/acssuschemeng.8b03454]
32. Gottardo M, Micolucci F, Bolzonella D, Uellendahl H, Pavan P. Pilot scale fermentation coupled with anaerobic digestion of food waste - effect of dynamic digestate recirculation. Renewable Energy 2017 Dec;114:455-463. [doi: 10.1016/j.renene.2017.07.047]
33. Estévez-Alonso Á, Pei R, van Loosdrecht MCM, Kleerebezem R, Werker A. Scaling-up microbial community-based polyhydroxyalkanoate production: status and challenges. Bioresour Technol 2021 May;327:124790. [doi: 10.1016/j.biortech.2021.124790] [Medline: 33582521]
34. Morgan-Sagastume F, Hjort M, Cirne D, et al. Integrated production of polyhydroxyalkanoates (PHAs) with municipal wastewater and sludge treatment at pilot scale. Bioresour Technol 2015 Apr;181:78-89. [doi: 10.1016/j.biortech.2015.01.046] [Medline: 25638407]
35. Valentino F, Morgan-Sagastume F, Campanari S, Villano M, Werker A, Majone M. Carbon recovery from wastewater through bioconversion into biodegradable polymers. N Biotechnol 2017 Jul 25;37(Pt A):9-23. [doi: 10.1016/j.nbt.2016.05.007] [Medline: 27288751]
36. Metcalf & Eddy, Inc, Tchobanoglous G, Stensel H, Tsuchihashi R, Burton F. Wastewater Engineering: Treatment, Disposal, and Reuse, 5th edition: McGraw-Hill Education; 2014.
37. Tamis J, Joosse BM, Loosdrecht MCMV, Kleerebezem R. High-rate volatile fatty acid (VFA) production by a granular sludge process at low pH. Biotechnol Bioeng 2015 Nov;112(11):2248-2255. [doi: 10.1002/bit.25640] [Medline: 25950759]
38. Karki R, Chuenchart W, Surendra KC, Sung S, Raskin L, Khanal SK. Anaerobic co-digestion of various organic wastes: kinetic modeling and synergistic impact evaluation. Bioresour Technol 2022 Jan;343:126063. [doi: 10.1016/j.biortech.2021.126063] [Medline: 34619321]
39. Moreno AD, Magdalena JA, Oliva JM, et al. Sequential bioethanol and methane production from municipal solid waste: an integrated biorefinery strategy towards cost-effectiveness. Process Safety Environ Protection 2021 Feb;146:424-431. [doi: 10.1016/j.psep.2020.09.022]

Abbreviations

BMP: bio-methane potential

CH_4: methane
CO_2: carbon dioxide
COD: chemical oxygen demand
FS/IN: feedstock/inoculum
H_2: hydrogen molecule
HRT: hydraulic retention time
kPa: kilopascal
MJ: megajoules
MWh: megawatt-hour
N-NH_4^+: ammonium
N_2: nitrogen molecule
NaOH: sodium hydroxide
O_2: oxygen molecule
OLR: organic loading rate
OMSW: organic municipal solid waste
P: phosphorous
P-PO_4^{3-}: phosphate
PE: population equivalent
PHA: polyhydroxyalkanoate
rpm: rounds per minute
SCOD: soluble chemical oxygen demand
SGP: specific biogas production
SMP: specific methane production
TS: total solids
v/v: volumetric basis
v/v %: maximum volumetric methane content
VFA: volatile fatty acid
VS: volatile solids
WS: waste sludge
WWTP: wastewater treatment plant

Edited by T Leung; submitted 01.07.23; peer-reviewed by D Elsalamony, Anonymous; revised version received 08.07.24; accepted 12.07.24; published 04.02.25.

Please cite as:
Borhany H
Converting Organic Municipal Solid Waste Into Volatile Fatty Acids and Biogas: Experimental Pilot and Batch Studies With Statistical Analysis
JMIRx Med 2025;6:e50458
URL: https://xmed.jmir.org/2025/1/e50458
doi:10.2196/50458

Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance

Masab Mansoor[1], BS, MBA, DBA; Andrew F Ibrahim[2], BS; David Grindem[3], DO; Asad Baig[4], MD

[1]Edward Via College of Osteopathic Medicine, 4408 Bon Aire Dr, Monroe, LA, United States

[2]Texas Tech University Health Sciences Center School of Medicine, Lubbock, TX, United States

[3]Mayo Clinic, Rochester, MN, United States

[4]Department of Radiology, Columbia University Medical Center, New York, NY, United States

Corresponding Author:
Masab Mansoor, BS, MBA, DBA
Edward Via College of Osteopathic Medicine, 4408 Bon Aire Dr, Monroe, LA, United States

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.09.24311777v1

Companion article: https://med.jmirx.org/2025/1/e73264

Companion article: https://med.jmirx.org/2025/1/e73258

Abstract

Background: Rural health care providers face unique challenges such as limited specialist access and high patient volumes, making accurate diagnostic support tools essential. Large language models like GPT-3 have demonstrated potential in clinical decision support but remain understudied in pediatric differential diagnosis.

Objective: This study aims to evaluate the diagnostic accuracy and reliability of a fine-tuned GPT-3 model compared to board-certified pediatricians in rural health care settings.

Methods: This multicenter retrospective cohort study analyzed 500 pediatric encounters (ages 0 - 18 years; n=261, 52.2% female) from rural health care organizations in Central Louisiana between January 2020 and December 2021. The GPT-3 model (DaVinci version) was fine-tuned using the OpenAI application programming interface and trained on 350 encounters, with 150 reserved for testing. Five board-certified pediatricians (mean experience: 12, SD 5.8 years) provided reference standard diagnoses. Model performance was assessed using accuracy, sensitivity, specificity, and subgroup analyses.

Results: The GPT-3 model achieved an accuracy of 87.3% (131/150 cases), sensitivity of 85% (95% CI 82% - 88%), and specificity of 90% (95% CI 87% - 93%), comparable to pediatricians' accuracy of 91.3% (137/150 cases; *P*=.47). Performance was consistent across age groups (0 - 5 years: 54/62, 87%; 6 - 12 years: 47/53, 89%; 13 - 18 years: 30/35, 86%) and common complaints (fever: 36/39, 92%; abdominal pain: 20/23, 87%). For rare diagnoses (n=20), accuracy was slightly lower (16/20, 80%) but comparable to pediatricians (17/20, 85%; *P*=.62).

Conclusions: This study demonstrates that a fine-tuned GPT-3 model can provide diagnostic support comparable to pediatricians, particularly for common presentations, in rural health care. Further validation in diverse populations is necessary before clinical implementation.

(JMIRx Med 2025;6:e65263) doi:10.2196/65263

KEYWORDS

natural language processing; NLP; machine learning; ML; artificial intelligence; language model; large language model; LLM; generative pretrained transformer; GPT; pediatrics

Introduction

The rapid advancement of artificial intelligence (AI) has led to the development of large language models (LLMs) that demonstrate sophisticated capabilities in understanding and analyzing human language [1]. Recent studies have shown promising applications of LLMs in health care, particularly in clinical decision support, medical knowledge synthesis, and diagnostic assistance [2-4]. However, their reliability and accuracy in specialized medical domains, especially pediatric care in resource-constrained settings, require thorough evaluation.

Differential diagnosis in pediatrics presents unique challenges that distinguish it from adult medicine. Young patients often cannot articulate their symptoms clearly, presentations can be atypical, and the range of potential diagnoses varies significantly with age. Recent systematic reviews have shown that diagnostic errors occur in "appreciable amounts" of pediatric encounters, with higher rates in rural and underserved areas [5]. These errors can lead to delayed treatment, inappropriate interventions, and potentially adverse outcomes.

The application of LLMs in clinical decision support has shown initial promise. Studies using GPT-3 and similar models have reported accuracies ranging from 75% to 85% in generating differential diagnoses for adult cases [6]. Notably, Steinberg et al [7] demonstrated that LLMs could achieve 82% accuracy in analyzing electronic health record (EHR) data for diagnostic support. However, pediatric applications remain underexplored, with limited studies specifically examining LLM performance in child and adolescent cases.

Rural health care settings face particular challenges that could benefit from LLM-based support tools. These areas often experience physician shortages, with providers managing high patient volumes and limited access to specialist consultation [8]. A survey of rural pediatric practices found that 52% of rural pediatricians report difficulty obtaining timely specialist input for complex cases [9]. Additionally, rural providers often work in isolation, managing a broad spectrum of conditions with fewer diagnostic resources compared to urban centers [10].

Previous evaluations of AI in pediatric diagnosis have largely focused on specific conditions or imaging-based applications rather than broad differential diagnosis. For instance, Wu et al [11] achieved 97.45% accuracy in pediatric otitis media interpretation using deep learning models, while other studies have demonstrated AI's effectiveness in detecting pediatric pneumonia from chest x-rays or identifying developmental disorders through automated screening tools. However, these models are often constrained by narrow diagnostic scopes, lack interpretability, and are not readily adaptable to general pediatric clinical reasoning.

Recent studies have begun to explore the application of LLMs in pediatric clinical settings. For example, Nian et al [12] found that ChatGPT and Google Gemini performed inadequately in providing recommendations for managing developmental dysplasia of the hip compared to expert guidelines, raising concerns about reliability in pediatric decision-making. Similarly, Wang et al [13] developed an LLM-based framework for pediatric obstructive sleep apnea management, highlighting the potential for specialized fine-tuning to improve diagnostic accuracy in specific pediatric conditions. Miyake et al [14] explored the role of AI-driven LLMs in pediatric surgery, emphasizing challenges related to real-time intraoperative decision support. Furthermore, Raza et al [15] investigated LLM applications in analyzing parental transcripts for children with congenital heart disease, demonstrating their potential role in augmenting thematic analysis in pediatric health care.

Despite these developments, comprehensive evaluations of LLMs in general pediatric differential diagnosis remain scarce. Many existing studies focus on narrow applications, lack real-world clinical validation, or fail to address age-specific nuances in pediatric presentations. Additionally, research on LLM utility in rural settings, where pediatricians may have limited access to specialist support, is particularly lacking. This study aims to bridge these gaps by systematically evaluating LLM performance in general pediatric differential diagnosis, with a focus on rural applicability and real-world clinical decision support.

The emergence of newer LLM architectures and their potential application in health care necessitates rigorous evaluation in real-world clinical settings [16]. While preliminary studies suggest promise, questions remain about their reliability, safety, and integration into clinical workflows [17]. Furthermore, the unique aspects of pediatric care—including age-specific disease presentations, developmental considerations, and the critical nature of early accurate diagnosis—require specific validation of these tools in pediatric populations [18].

This study addresses these knowledge gaps by evaluating the performance of a fine-tuned GPT-3 model in generating pediatric differential diagnoses within rural health care settings. By comparing the model's performance with that of experienced pediatricians across various age groups and presenting complaints, we aim to assess its potential as a clinical decision support tool. The findings could inform the development of AI-assisted diagnostic tools specifically tailored to the needs of rural pediatric health care providers.

Methods

Study Design and Setting

This multicenter retrospective cohort study was conducted in collaboration with a rural pediatric health care organization in Central Louisiana. The organization provides primary care to approximately 15,000 pediatric patients. The study analyzed patient data collected between January 2020 and December 2021. The overall workflow of the study is illustrated in Figure 1, encompassing data collection through model evaluation.

Figure 1. Workflow schematic showing the process of data collection, preprocessing, model training, and evaluation. The pipeline includes data splitting (70% training, 30% testing), GPT-3 fine-tuning, and comprehensive performance evaluation including subgroup analyses.

Ethical Considerations

Ethics approval was obtained from the Mansoor Pediatrics Ethics Committee (approval MP-2023‐017), and the study adhered to the principles of the Declaration of Helsinki. The study used retrospective, deidentified patient data and was exempt from informed consent requirements. Data were anonymized to ensure compliance with Health Insurance Portability and Accountability Act (HIPAA) regulations. No identifying information was accessible to researchers. No compensation was provided to participants as the study relied on existing retrospective data. For secondary analyses using deidentified data, the original consent obtained at the time of patient care covered the use of the data for research purposes.

Participants and Data Collection

A total of 500 pediatric patient encounters were included based on the following criteria:

- Inclusion criteria: Patients aged 0‐18 years with a documented chief complaint and pediatrician-generated differential diagnosis
- Exclusion criteria: Encounters with incomplete or inconsistent data

Anonymized data, including patient age, sex, chief complaint, presenting symptoms, medical history, and pediatrician-generated differential diagnoses, were extracted from the EHR system. Two independent researchers manually reviewed the data to ensure accuracy and consistency. No missing data were present in the final dataset. Demographic information, including racial and ethnic background, was not collected as part of this dataset. This omission limits the ability to assess potential biases in model performance across racial or ethnic groups, which is an important consideration for future research.

Five board-certified pediatricians (mean experience: 12, SD 5.8, range 5‐20 years) participated in the study as reference standard providers. Pediatricians were recruited from the participating health care organization based on their availability and experience in rural pediatrics.

Data Preprocessing

For each patient encounter, the chief complaint, presenting symptoms, and relevant medical history were concatenated into a single text string. Identifying information was removed to ensure privacy. Medical terms were standardized using a medical dictionary, and data were formatted for compatibility with the GPT-3 model.

Model Training and Fine-Tuning

The GPT-3 model (DaVinci version) was fine-tuned using the OpenAI application programming interface. The dataset was randomly split into a training set (n=350, 70%) and a testing set (n=150, 30%). The model was trained to generate up to five differential diagnoses for each input case. The study used retrospective data that included pediatrician-generated differential diagnoses documented during actual clinical encounters. No pediatricians were prospectively instructed to generate differential diagnoses specifically for this study. The same format of up to 5 differential diagnoses was used for standardization when processing both the historical physician documentation and the GPT-3 outputs. Fine-tuning parameters included 10 epochs, a batch size of 4, and a learning rate of 1e-5. The fine-tuning process aimed to optimize the model's ability to generate accurate and relevant differential diagnoses based on the input data. These details are visible in Multimedia Appendix 1.

GPT-3 (DaVinci version) was selected for this study because it was the most advanced version of the GPT model available at the time of data collection and model fine-tuning. Subsequent versions, such as GPT-3.5 and GPT-4, were released after the study period and were therefore not considered. Future work could explore the performance of these newer models in similar settings to assess potential improvements in diagnostic accuracy.

Evaluation Metrics

The model's performance was evaluated using the following metrics (Table 1):

- Accuracy: Proportion of correct predictions (true positives and true negatives) relative to total cases
- Sensitivity (recall): Proportion of actual positive diagnoses correctly identified by the model
- Specificity: Proportion of actual negative diagnoses correctly excluded by the model
- Precision: Proportion of positive predictions that were correct
- F_1-score: Harmonic mean of precision and sensitivity

In addition to these metrics, subgroup analyses were conducted by age group (0‐5, 6‐12, and 13‐18 years) and chief complaints (eg, fever, abdominal pain).

Table . Testing set evaluation metrics for analysis of the fine-tuned GPT-3 model, including formulas and values of the evaluation metrics for the GPT-3 model.

Metric	Formula	Description
Sensitivity (recall)	TP[a,b]/(TP + FN[c,d])	The proportion of actual positive diagnoses that were correctly identified by the model
Specificity	TN[e,f]/(TN + FP[g,h]) 0.90	The proportion of actual negative diagnoses that were correctly identified by the model
Precision	TP/(TP + FP)	The proportion of the model's positive predictions that were actual positive diagnoses
F_1-score	2 * (precision * sensitivity)/(precision + sensitivity)	The harmonic mean of precision and sensitivity, providing a balanced measure of the model's performance
Accuracy	(TP + TN)/(TP + TN + FP + FN)	The overall proportion of correct predictions made by the model

[a]TP: true positive.

[b]Cases where the model correctly predicted a positive diagnosis.

[c]FN: false negative.

[d]Cases where the model incorrectly predicted a negative diagnosis.

[e]TN: true negative.

[f]Cases where the model correctly predicted a negative diagnosis.

[g]FP: false positive.

[h]Cases where the model incorrectly predicted a positive diagnosis.

Statistical Analysis

Descriptive statistics were used to summarize patient demographics and model performance. χ^2 tests were used for categorical variables, and independent 2-tailed *t* tests were used for continuous variables. Statistical significance was set at $P<.05$. Data normality was assessed using the Kolmogorov-Smirnov test before statistical analysis. Our outcome metrics (accuracy, sensitivity, specificity) were found to follow a normal distribution ($P>.05$), supporting our use of parametric statistical methods including *t* tests for comparisons between groups. For nonnormally distributed variables, nonparametric alternatives (Mann-Whitney *U* test) were applied.

χ^2 tests were chosen for categorical variables due to their robustness in comparing proportions across groups. Independent *t* tests were selected for continuous variables after confirming normality of distribution. The choice of metrics (accuracy, sensitivity, specificity) aligns with standard diagnostic evaluation frameworks in health care AI validation studies. Subgroup analyses were performed to assess model performance consistency across demographics and clinical presentations, which is essential for evaluating potential biases in model predictions.

Power analysis indicated that a sample size of 500 would provide 80% power to detect a 10% difference in accuracy between the GPT-3 model and pediatricians, assuming a pediatrician accuracy of 90%. This calculation accounted for the expected distribution of common and rare diagnoses in our pediatric population, with consideration for potential subgroup analyses across different age groups and chief complaints.

Software and Tools

The statistical analysis was conducted using Python 3.8 (Python Software Foundation) [19] with the scikit-learn library [20] for model evaluation and SPSS Statistics version 29 (IBM Corp) for additional analysis [21]. The OpenAI application programming interface was used for model fine-tuning and prediction generation [22]. Software and scripts used in this study are available upon request for reproducibility.

Results

Dataset Characteristics

A total of 500 pediatric patient encounters were included, with 350 (70%) cases in the training set and 150 (30%) cases in the testing set. The mean age of patients was 7.5 (SD 5.2) years, and 52.2% (n=261) of participants were female. The most common chief complaints were fever (n=130, 26%), cough (n=98, 19.6%), abdominal pain (n=73, 14.6%), and rash (n=49, 9.8%). The distribution of age, sex, and chief complaint was similar between the training and testing sets (Table 2).

Table . Demographics and dataset characteristics.

Characteristic	Total (N=500)	Training set (n=350)	Testing set (n=150)	*P* value
Age (years), mean (SD)	7.5 (5.2)	7.4 (5.1)	7.7 (5.3)	.56[a]
Sex, n (%)				.82[b]
Female	261 (52.2)	184 (52.6)	77 (51.3)	
Male	239 (47.8)	166 (47.4)	73 (48.7)	
Chief complaint, n (%)				.93[b]
Fever	130 (26.0)	91 (26.0)	39 (26.0)	
Cough	98 (19.6)	70 (20.0)	28 (18.7)	
Abdominal pain	73 (14.6)	50 (14.3)	23 (15.3)	
Rash	49 (9.8)	34 (9.7)	15 (10.0)	
Other	150 (30.0)	105 (30.0)	45 (30.0)	
Rare diagnoses, n (%)	20 (4.0)	14 (4.0)	6 (4.0)	>.99

[a]*P* value calculated using independent 2-tailed *t* test.

[b]*P* value calculated using χ^2 test.

Model Performance

The fine-tuned GPT-3 model achieved high accuracy in generating differential diagnoses on the testing set. Key performance metrics are as follows:

- Accuracy: 87.3% (131/150 cases)
- Sensitivity (recall): 85% (95% CI 82% - 88%)
- Specificity: 90% (95% CI 87% - 93%)
- Precision: 89% (95% CI 86% - 92%)
- F_1-score: 0.87

The model correctly identified 128 positive diagnoses and excluded 334 negative diagnoses, with 16 false positives and 22 false negatives.

Subgroup Analysis

Performance across age groups and common chief complaints are summarized in Tables 2 and 3. The model's accuracy was consistent across age groups:

- 0 - 5 years: 87% (54/62 cases)
- 6 - 12 years: 89% (47/53 cases)
- 13 - 18 years: 86% (30/35 cases)

Table . Model performance by common chief complaints.

Chief complaint	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Precision (95% CI)	F_1-score (95% CI)
Fever (n=39)	0.92 (0.88-0.96)	0.90 (0.85-0.95)	0.93 (0.90-0.96)	0.92 (0.87-0.97)	0.91 (0.86-0.96)
Cough (n=28)	0.89 (0.82-0.94)	0.85 (0.79-0.91)	0.90 (0.84-0.92)	0.89 (0.83-0.95)	0.87 (0.81-0.93)
Abdominal pain (n=23)	0.87 (0.78-0.92)	0.82 (0.75-0.89)	0.87 (0.83-0.90)	0.86 (0.79-0.93)	0.84 (0.77-0.91)
Rash (n=15)	0.93 (0.83-0.97)	0.88 (0.80-0.96)	0.91(0.88-0.94)	0.90 (0.92-0.98)	0.89 (81-0.97)

Similarly, the model demonstrated robust performance for common chief complaints:

- Fever: 92% (36/39 cases) accuracy
- Cough: 89% (25/28) accuracy
- Abdominal pain: 87% (20/23) accuracy
- Rash: 93% (14/15) accuracy

Subgroup analyses by age group and chief complaints revealed consistent performance, indicating the model's ability to generalize across varying pediatric presentations. However, the slight performance drop in complex and rare cases underscores the importance of targeted training datasets for improving diagnostic accuracy in these subgroups. For rare or complex diagnoses (n=20), the model achieved an accuracy of 80% (16/20 cases), slightly lower than the overall accuracy but comparable to pediatricians (17/20, 85% of cases; *P*=.62).

Comparison With Pediatricians

The model's performance was comparable to that of the 5 participating board-certified pediatricians. Pediatricians achieved an accuracy of 91.3% (137/150 cases), with a sensitivity of 92% (95% CI 91%-94%) and specificity of 88% (95% CI 84%-90%). Differences in sensitivity (*P*=.08) and specificity (*P*=.57) between the model and pediatricians were not statistically significant.

Statistical Analysis

χ^2 tests indicated no significant differences between the GPT-3 model and pediatricians for accuracy, sensitivity, or specificity. Subgroup analyses confirmed consistent performance across age groups and common chief complaints, with no significant performance disparities.

Tables

Table 1 provides a detailed breakdown of the evaluation metrics. Table 4 shows the performance of the model by age group, while Table 3 summarizes performance by chief complaints.

Table . Model performance by age group.

Age group (years)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Precision (95% CI)	F_1-score (95% CI)
Overall (n=150)	0.85 (0.81-0.89)	0.90 (0.87-0.93)	0.89 (0.86-0.92)	0.87 (0.83-0.91)	0.88 (0.85-0.91)
0 - 5 (n=62)	0.87 (0.82-0.92)	0.84 (0.79-0.89)	0.89 (0.85-0.93)	0.88 (0.83-0.93)	0.86 (0.81-0.91)
6 - 12 (n=53)	0.89 (0.84-0.94)	0.86 (0.81-0.91)	0.91 (0.87-0.95)	0.90 (0.85-0.95)	0.88 (0.83-0.93)
13 - 18 (n=35)	0.86 (0.80-0.92)	0.83 (0.77-0.89)	0.88 (0.83-0.93)	0.87 (0.81-0.93)	0.85 (0.79-0.91)

Discussion

Principal Findings

This study evaluated the diagnostic performance of a fine-tuned GPT-3 model in generating pediatric differential diagnoses in rural health care settings. The model achieved an accuracy of 87%, which was comparable to board-certified pediatricians' accuracy of 91%. Performance was consistent across age groups and common chief complaints, underscoring the model's potential as a reliable clinical decision support tool. While the model demonstrated lower accuracy for rare or complex cases (80%), its performance remained comparable to that of pediatricians (85%). These findings suggest that LLMs could enhance diagnostic accuracy and support providers in underserved regions, particularly for routine presentations.

Comparison to Prior Work

Our findings align with prior studies demonstrating the potential of LLMs in clinical decision support. For example, Steinberg et al [7] reported 82% accuracy in adult diagnostic support using LLMs, while Wu et al [11] achieved 97.45% accuracy in pediatric otitis media interpretation with deep learning models. This study extends these findings by focusing on general pediatric differential diagnosis, an area with limited prior research. Unlike previous studies that primarily examined urban or hospital-based datasets, our work highlights the utility of LLMs in resource-constrained rural environments, addressing a critical gap in the literature.

Strengths and Limitations

This study has several strengths. First, the use of real-world data from rural health care settings enhances the generalizability of findings to similar environments. Second, the inclusion of subgroup analyses provides insights into the model's performance across diverse age groups and chief complaints. Third, the comparative evaluation with experienced pediatricians underscores the model's clinical relevance.

Another of the key strengths of this study lies in its real-world applicability, particularly for rural health care settings where resources are limited and access to specialists is often constrained. By leveraging existing EHR data and evaluating the model's performance on common and rare pediatric conditions, this research provides a practical framework for integrating AI tools into primary care workflows. The consistent accuracy demonstrated across age groups and chief complaints highlights the potential of GPT-3 to serve as a valuable diagnostic support system for providers in underserved areas. However, implementing such tools in real-world clinical settings will require addressing infrastructure challenges, including internet connectivity and provider training. Despite these challenges, the findings underscore the feasibility of deploying AI systems to enhance diagnostic accuracy and reduce disparities in health care delivery, particularly in environments with high patient volumes and limited specialist availability.

However, there are notable limitations:

- Sample size and diversity: The sample size of 500 encounters, while informative, may not fully capture the diversity of the broader pediatric population. This limitation is particularly relevant in diverse health care settings, where factors such as demographic variability, socioeconomic status, and health care access can influence diagnostic patterns. Prior studies have demonstrated that models trained on limited datasets often fail to generalize across different populations, highlighting the need for larger, multi-institutional datasets to improve validity and applicability [17]. Additionally, our study used data from a single rural health care organization, which may limit the external validity of our findings. Similar studies have shown that AI-based diagnostic models exhibit performance degradation when applied to new patient populations due to variations in disease prevalence, clinical workflows, and physician documentation styles [18]. For instance, Steinberg et al [7] found that an LLM trained on one hospital's EHRs experienced a 15% drop in accuracy when tested on data from a different institution. These findings emphasize the need for external validation. Future research should prioritize expanding the sample size through multicenter collaborations, incorporating data from health care centers with diverse patient demographics to enhance generalizability and robustness. Similar initiatives have demonstrated improved AI model performance when trained on heterogeneous datasets, such as the multi-institutional validation study by Rajkomar et al [2], which improved diagnostic accuracy across multiple health care networks.
- Retrospective design: The use of retrospective data limits the ability to assess the model's impact on clinical workflows or patient outcomes. Prospective clinical trials are needed to evaluate these aspects.
- Cross-validation: A key limitation of this study is the lack of cross-validation across different health care organizations. Evidence suggests that AI-based diagnostic models frequently underperform when tested on external

datasets due to variations in clinical documentation, patient demographics, and institutional practices. For example, a systematic review of AI applications in health care found that models trained on single-center data exhibited an average 12% - 20% decrease in performance when applied to external datasets [17]. Steinberg et al [7] also demonstrated that LLMs trained on EHRs from one hospital struggled to maintain accuracy when exposed to unseen patient populations, emphasizing the importance of cross-validation. Furthermore, ChatGPT-based diagnostic models have shown variability in reliability across different patient demographics, particularly when applied to pediatric populations with rare conditions [12]. To ensure reproducibility, future studies should incorporate external validation using data from multiple institutions, including urban, suburban, and rural health care settings. By validating performance across diverse patient populations, we can assess the model's reliability in real-world clinical environments and mitigate the risks associated with dataset bias. This approach aligns with recommendations from previous research advocating for multicenter validation to improve AI model robustness [18].

- Rare diagnoses: The model's lower accuracy for rare or complex cases highlights the need for further fine-tuning and testing in these areas. Future fine-tuning efforts could incorporate domain-specific datasets, such as rare pediatric conditions or uncommon presentations, to enhance the model's diagnostic accuracy for less frequently encountered cases. For example, fine-tuning could focus on rare pediatric conditions such as Kawasaki disease or metabolic disorders, which often present atypically and are prone to diagnostic errors. Collaborations with specialist clinics could help build robust datasets for such conditions.
- GPT-3 versus newer models: Another limitation is the use of GPT-3 instead of its newer iterations, such as GPT-3.5 or GPT-4, which were released after the completion of this study. While GPT-3 demonstrated strong diagnostic performance, future studies should evaluate whether these more advanced models can further enhance accuracy, particularly for rare or complex cases. Specifically, GPT-3.5 and GPT-4 feature enhanced contextual understanding and larger training corpora [23], which may improve their ability to identify nuanced patterns in rare pediatric diagnoses. Additionally, these models may mitigate hallucination risks and offer better attribution of sources, which are critical for clinical applications. Comparative evaluations in similar rural health care settings would provide insights into their incremental benefits over GPT-3.

Practical Implications

Integrating LLMs like GPT-3 into rural health care settings could address critical challenges such as physician shortages, high patient volumes, and limited specialist access. These tools can provide rapid accurate diagnostic support, reducing diagnostic errors and improving patient outcomes [24]. However, practical barriers to implementation, including infrastructure requirements (eg, reliable internet and electricity) and provider training, must be addressed [25].

Reliance on AI systems poses risks, including overreliance by less experienced providers and challenges in managing incomplete or inconsistent input data [26]. Training programs should ensure health care providers understand the limitations of AI tools and develop strategies for validating AI-generated outputs. Establishing clear guidelines for AI use in clinical settings will further ensure patient safety and ethical application. To address concerns about hallucinations—instances where the model generates inaccurate or fabricated information—health care providers must verify AI-generated outputs against clinical guidelines and existing evidence. Integrating feedback mechanisms, where physicians can flag inaccuracies, may also help refine model behavior over time [27].

Additionally, fostering trust in AI tools among providers and patients will be essential for successful adoption [28]. Additionally, parental concerns regarding deferring diagnostic decisions to AI systems must be addressed to build trust and acceptance. Efforts to educate families about AI's role as a supplementary decision-making tool rather than a replacement for physician judgment are essential. Furthermore, rural health care facilities may face challenges in implementing AI solutions due to limited infrastructure, such as inconsistent internet access, power supply, and provider training [29]. These challenges may also include the cost of deploying and maintaining AI systems, as well as the need for ongoing technical support. Policy makers and health care administrators should explore subsidized programs or partnerships with technology providers to ensure equitable access to AI tools in resource-limited settings. Addressing these barriers will be crucial for ensuring successful adoption and integration into clinical workflows.

Future Directions

Future research should focus on the following:

- The findings should be validated in larger, more diverse populations across multiple health care settings.
- The diagnostic capabilities of more advanced models, such as GPT-3.5 or GPT-4, should be assessed to determine whether recent improvements in language model architecture further enhance diagnostic accuracy.
- The impact of LLM integration on patient outcomes, provider satisfaction, and workflow efficiency in prospective clinical trials should be assessed.
- User-friendly interfaces should be developed to facilitate adoption by providers with varying levels of technological expertise, and training programs tailored to rural health care providers should be developed to familiarize them with AI tools and address potential apprehensions about using such systems. These programs should emphasize the complementary nature of AI in clinical workflows rather than its replacement of human judgment.
- Ethical concerns, including data privacy, informed consent, and model transparency, should be addressed to ensure responsible use in clinical practice.
- In addition to traditional evaluation metrics, future studies should assess language generation issues such as hallucinations—instances where the model produces false or unsupported information—and attribution of responses to reliable sources.

These factors are critical for ensuring the safety and reliability of AI applications in clinical decision-making. Natural language processing metrics like Recall-Oriented Understudy for Gisting Evaluation (ROUGE) and bilingual evaluation understudy (BLEU) may be used to evaluate output quality, while further human review of generated responses could assess alignment with established clinical guidelines.

Conclusions

This study highlights the potential of GPT-3, a fine-tuned LLM, as a clinical decision support tool for pediatric differential diagnosis in rural health care settings. The model achieved diagnostic accuracy comparable to that of board-certified pediatricians, demonstrating robust performance across age groups and common presenting complaints. These findings suggest that LLMs could serve as valuable tools for addressing the unique challenges faced by rural health care providers, such as limited access to specialists and high patient volumes.

However, this work also underscores the need for further validation. Future research should focus on evaluating the model's performance in larger, diverse populations and real-world clinical settings. Ethical considerations, including data privacy and model transparency, must be prioritized to ensure responsible implementation. Another ethical consideration is the potential for AI models to exacerbate existing health disparities if their development does not account for diverse populations. Rigorous testing in underrepresented groups and ongoing audits for bias are critical steps to ensure fairness and equity in AI-driven health care applications. By addressing these challenges, LLMs like GPT-3 have the potential to enhance diagnostic accuracy, reduce disparities in access to care, and improve outcomes for pediatric patients in underserved regions.

While this study represents a step toward integrating AI into rural health care, its findings underscore the need for iterative improvements and cross-disciplinary collaboration to refine these tools. Partnerships between AI developers, clinicians, and health care administrators will be crucial in ensuring that AI solutions are both effective and accessible.

This study serves as a step in bridging the gap between AI innovation and practical health care applications, paving the way for future advancements in clinical decision support systems tailored to the needs of rural health care environments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Technical appendix: GPT-3 Model specifications and implementation details.
[DOCX File, 24 KB - xmed_v6i1e65263_app1.docx]

References

1. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems 33 (NeurIPS 2020): Neural Information Processing Systems Foundation, Inc; 2020:1877-1901.
2. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1:18. [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
3. Liu Y, Ott M, Goyal N, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online on Jul 26, 2019. [doi: 10.48550/ARXIV.1907.11692]
4. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 4;47(1):33. [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
5. Rinke ML, Singh H, Heo M, et al. Diagnostic errors in primary care pediatrics: Project RedDE. Acad Pediatr 2018 Mar;18(2):220-227. [doi: 10.1016/j.acap.2017.08.005] [Medline: 28804050]
6. Hirosawa T, Kawamura R, Harada Y, et al. ChatGPT-generated differential diagnosis lists for complex case-derived clinical vignettes: diagnostic accuracy evaluation. JMIR Med Inform 2023 Oct 9;11(1):e48808. [doi: 10.2196/48808] [Medline: 37812468]
7. Steinberg E, Jung K, Fries JA, Corbin CK, Pfohl SR, Shah NH. Language models are an effective representation learning technique for electronic health record data. J Biomed Inform 2021 Jan;113:103637. [doi: 10.1016/j.jbi.2020.103637] [Medline: 33290879]
8. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan;79(1-2):169-176. [doi: 10.1038/pr.2015.192] [Medline: 26466080]
9. Pletcher BA, Rimsza ME, Cull WL, Shipman SA, Shugerman RP, O'Connor KG. Primary care pediatricians' satisfaction with subspecialty care, perceived supply, and barriers to care. J Pediatr 2010 Jun;156(6):1011-1015. [doi: 10.1016/j.jpeds.2009.12.032] [Medline: 20227727]
10. Chipp C, Dewane S, Brems C, Johnson ME, Warner TD, Roberts LW. "If only someone had told me…": lessons from rural providers. J Rural Health 2011;27(1):122-130. [doi: 10.1111/j.1748-0361.2010.00314.x] [Medline: 21204979]
11. Wu Z, Lin Z, Li L, et al. Deep learning for classification of pediatric otitis media. Laryngoscope 2021 Jul;131(7):E2344-E2351. [doi: 10.1002/lary.29302] [Medline: 33369754]

12. Nian PP, Umesh A, Jones RH, et al. ChatGPT and Google Gemini are clinically inadequate in providing recommendations on management of developmental dysplasia of the hip compared to American Academy of Orthopaedic Surgeons Clinical Practice Guidelines. J Pediatr Soc North Am 2025 Feb;10:100135. [doi: 10.1016/j.jposna.2024.100135]
13. Wang Q, Cao Z, Mao Q, et al. OSAer: a specialized LLM framework for pediatric obstructive sleep apnea management. SSRN. Preprint posted online on Jan 27, 2025. [doi: 10.2139/ssrn.5109772]
14. Miyake Y, Retrosi G, Keijzer R. Artificial intelligence and pediatric surgery: where are we? Pediatr Surg Int 2024 Dec 3;41(1):19. [doi: 10.1007/s00383-024-05921-8] [Medline: 39625492]
15. Raza MZ, Xu J, Lim T, et al. LLM-TA: an LLM-enhanced thematic analysis pipeline for transcripts from parents of children with congenital heart disease. arXiv. Preprint posted online on Feb 3, 2025. [doi: 10.48550/arXiv.2502.01620]
16. Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. JAMA 2025 Jan 28;333(4):319-328. [doi: 10.1001/jama.2024.21700] [Medline: 39405325]
17. Bedi S, Liu Y, Orr-Ewing L, et al. A systematic review of testing and evaluation of healthcare applications of large language models (LLMs). medRxiv. Preprint posted online on Apr 16, 2024. [doi: 10.1101/2024.04.15.24305869]
18. Rader B, Hswen Y, Brownstein JS. Further reflections on the use of large language models in pediatrics. JAMA Pediatr 2024 Jun 1;178(6):628-629. [doi: 10.1001/jamapediatrics.2024.0729] [Medline: 38683628]
19. Phillips D. Python 3 Object-Oriented Programming: Build Robust and Maintainable Software With Object-Oriented Design Patterns in Python 38: Packt Publishing Ltd; 2018.
20. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Machine Learning Res 2011 Oct;12:2825-2830 [FREE Full text]
21. Meyers LS, Gamst GC, Guarino AJ. Performing Data Analysis Using IBM SPSS: John Wiley & Sons; 2013.
22. Kublik S, Saboo S. GPT-3: The Ultimate Guide to Building NLP Products With OpenAI API: Packt Publishing; 2023.
23. Šimsová J. Examining cognitive abilities and multilingual performance of large language models: a comparative analysis of GPT-3 and GPT-4 [thesis]. : Charles University; 2024 URL: https://dspace.cuni.cz/handle/20.500.11956/195021 [accessed 2025-03-11]
24. Guo J, Li B. The application of medical artificial intelligence technology in rural areas of developing countries. Health Equity 2018 Aug 1;2(1):174-181. [doi: 10.1089/heq.2018.0037] [Medline: 30283865]
25. Khan F, Driessen J. Bridging the telemedicine infrastructure gap: implications for long-term care in rural America. Public Policy Aging Rep 2018 Nov 2;28(3):80-84. [doi: 10.1093/ppar/pry027]
26. Muley A, Muzumdar P, Kurian G, Basyal GP. Risk of AI in healthcare: a comprehensive literature review and study framework. Asian J Med Health 2023 Aug 28;21(10):276-291. [doi: 10.9734/ajmah/2023/v21i10903]
27. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 1;27(3):491-497. [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
28. Sîrbu CL, Mercioni MA. Fostering trust in AI-driven healthcare: a brief review of ethical and practical considerations. Presented at: 2024 International Symposium on Electronics and Telecommunications (ISETC); Nov 7-8, 2024; Timisoara, Romania. [doi: 10.1109/ISETC63109.2024.10797264]
29. Olugboja A, Agbakwuru EM. Bridging healthcare disparities in rural areas of developing countries: leveraging artificial intelligence for equitable access. Presented at: 2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA); Feb 1-2, 2024; Victoria, Seychelles. [doi: 10.1109/ACDSA59508.2024.10467443]

Abbreviations

AI: artificial intelligence
BLEU: bilingual evaluation understudy
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
LLM: large language model
ROUGE: Recall-Oriented Understudy for Gisting Evaluation

Edited by A Schwartz; submitted 10.08.24; peer-reviewed by D Saderi, G Bender, T Olatoye, A Rahgozar, U Kumar Chalwadi, E Nwanaforo, P Hassan Ilegbusi, S Sakilay, M Collier; revised version received 24.02.25; accepted 28.02.25; published 19.03.25.

Please cite as:
Mansoor M, Ibrahim AF, Grindem D, Baig A
Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance
JMIRx Med 2025;6:e65263
URL: https://xmed.jmir.org/2025/1/e65263
doi:10.2196/65263

Peer Review of “State Anxiety Biomarker Discovery: Electrooculography and Electrodermal Activity in Stress Monitoring (Preprint)”

Daniela Saderi[1]; Shailee Rasania[2]; Toba Olatoye[3]; Simon Muhindi Savai; Randa Salah Gomaa Mahmoud[4]; Vasco Medeiros; Mitchell Collier

[1]PREreview, Portland, OR, United States

[2]Inflammatix, Inc, King of Prussia, PA, United States

[3]Kwara State Teaching Service Commission, Ilorin, Nigeria

[4]Zagazig University, Zagazig, Egypt

Related Article:

Companion article: https://arxiv.org/abs/2411.17935v1

(JMIRx Med 2025;6:e72093) doi:10.2196/72093

KEYWORDS

stress; biomarker discovery; EOG; EEG; medical informatics; electrooculography; electroencephalography

This is a peer-review report for the preprint “State Anxiety Biomarker Discovery: Electrooculography and Electrodermal Activity in Stress Monitoring.”

This review is the result of a virtual collaborative live review discussion organized and hosted by PREreview and JMIR Publications on January 16, 2025. The discussion was joined by 16 people: 2 facilitators, 1 member of the JMIR Publications team, and 13 live review participants, including 3 who agreed to be named but have not contributed to composing this review into its final form: Uday Kumar Chalwadi, Killivalavan Solai, and Prasakthi Venkatesan. The authors of this review have dedicated additional asynchronous time over the course of 2 weeks to help compose this final report using the notes from the live review. We thank all participants who contributed to the discussion and made it possible for us to provide feedback on this preprint.

Summary

Anxiety, particularly state anxiety (s-anxiety), is increasingly recognized as a health concern linked to mental and physical issues, including adverse cardiovascular and long-term health outcomes. This study [1] leverages noninvasive wearable technology to identify interpretable biomarkers resulting from s-anxiety using electrooculography (EOG) and electrodermal activity (EDA). Two datasets were developed: BLINKEO, focusing on blink-related EOG features, and EMOCOLD, analyzing EOG and EDA responses during a cold pressor test. The authors then used both datasets and applied statistical analysis (eg, F_1-scoring, Shapley Additive Explanations [SHAP] analysis) to identify biomarkers of anxiety. Results revealed that using EOG data (blink duration, peak height, and opening integral) in tandem with EDA data (mean signal, permutation, entropy, and Hjorth activity) led to the identification of novel biomarkers that reveal nuanced emotional and stress responses. Moreover, it was found that SHAP analysis can more accurately determine which features are relevant to enhancing model performance. The findings highlight the potential of combining EOG and EDA biomarker data to create robust real-time models for anxiety detection. Combinations of physiological features (as sets) were more effective as measures of stress response than individual features alone. This research underscores the transformative role of noninvasive wearable technology in personalized mental health monitoring and intervention strategies.

List of Major Concerns and Feedback

Concerns With Methods

- It would be helpful to document the name of the device and manufacturer used to record the EOG. This would help other researchers who may want to reproduce the results.
- Similarly, it would be helpful to add additional details about the cold pressor test methods. For example, was a commercially available circulating water bath used to maintain a constant water temperature? Was the temperature of the subject’s hand monitored? The details of the cold stressor test (the water temperature, the period of immersion, and the cutoff point) should be added for the sake of clarity, transparency, and reproducibility. Past studies using these metrics should also be referenced for details (eg, [2]). These methodological details may also be added in the form of a figure to add clarity to the experimental setup.
- To better understand the individual response to the cold challenge before participating in the actual experiment, it is advised that the manuscript states what type of participant testing was or was not adopted in the cold pressor testing experiment. For example, what were the tolerance times?

Were there any gender differences? If any pretesting data were collected, analyzing them and presenting them as results would add clarity to the results.
- It is unclear if the 65 repeating blinking trials and the 19 no-blinking trials were collected from the same individual or from different individuals. Please clarify.
- No signal voltage/electrical records for EDA were found in the manuscript. Is this intentional? Please consider adding this information.
- It would be important to add details of ordinal variables present in the Positive and Negative Affect Schedule and the State-Trait Anxiety Inventory (STAI-State), and clearly state their function and use in Supplementary Table 2.

Concerns With Analysis

- F_1-scores that were mentioned in the text (87.34% and 79.99%) are not present within the figures. Moreover, an F_1-score is an integer value from 0 to 1, taking precision and recall into account, and is not often expressed as a percentage.
- Figure 1c has two separate graphs; it should be captioned as 1c and 1d. What do both these graphs portray? The second graph for 1c is missing titles for the x- and y-axes—the current assumption is that they are the same as the first graph.
- Table 1 lacks a legend and is shown as panel a of Table 2. Please check how the tables are referenced in the text to make sure they reference the right one.
- The captions of the figures should have statistical information when relevant. For example, in Figure 3, the caption should include a description of what data were plotted and the meaning of the graph. Presumably plotting medians, quartiles, and SDs? Also, please report n values.

Concerns With Ethics

- It is not clear what the ethical statement at the end of the manuscript, which states that the study was exempt from review board approval, means. That statement should be revised for clarification. In addition, details regarding whether or not institutional review board approval was obtained, whether the study involved consenting participants and used humans, how the data were collected and used, how the data were handled to protect the privacy of study participants, and any other ethical procedures that were followed to protect subjects from any harm due to participation in the study should be added.

List of Minor Concerns and Feedback

Minor Concerns With Methods

- Please document whether the data were taken from each subject only once or whether data were obtained several times from a subject.
- Referring to the line “To focus on blink-like events, we applied criteria based on established blink characteristics,” the criteria used to establish blink characteristics should be cited, if not already given.
- SHAP analysis was performed on combinations of 5 features. Please clarify on what basis these 5 features were chosen (out of 15 of EDG and 33 of EOG).

Minor Concerns With Analysis and Presentation

- Page 10, Electrooculography (EOG) Signal Segmentation section: the authors mentioned that they extracted 33 features; however, Supplementary 4 mentioned 35 feature definitions. Please revise and correct.
- In Figure 3, please put “STAI-State survey score” on the y-axis for clarification rather than just “Scores.” In addition to box and whiskers plots, adding column graphs for positive affectivity, negative affectivity, and s-anxiety might be beneficial to more clearly express the SD present within the data.
- It would be beneficial to graphically display the F_1-scores that were collected across the study.
- The figures are quite small, which makes readability a little difficult. Please make the text larger to improve readability and accessibility.
- The Figure 1a description states, “The red dotted lines indicate the center of the peak…,” but these appear to be gray.

Suggestions

- Consider the inclusion of a Limitations section in this manuscript to better discuss potential limitations due to the skewness in male and female participants, data curation, applied methodologies, and other limitations of the study.
- A figure showing the trial structure would be very useful to understand how the data were collected.

References

- In the third paragraph of the Introduction, adding a reference to other techniques used to provoke anxiety, including the reduced EDA response in depressed patients, and the conflicting studies could be helpful to the readers.
- In the Introduction, fourth paragraph, the reference “Schachter and Singer” is not present in the References. Is this the wrong reference, or it just needs to be added to the list?
- In the Introduction, third page, third paragraph, it is advised to add references to document the reduced EDA response in depressed patients and the conflicting studies.
- In the Methods, please cite sources for the Butterworth filter (page 5), the Savitzky-Golay filter (page 5), and all other analyses.
- Reference 2: Include full citation with a link.
- Reference 3: It is advised to correct the article name to “APA 2023 Stress in America Topline Data.”
- Reference 4: The correct citation should be “Kazanskiy NL., Khonina S.N., Butt M.A. A review on flexible wearables—Recent developments in non-invasive continuous health monitoring. Sens. Actuators A Phys. 2024;366:114993. doi: 10.1016/j.sna.2023.114993.”
- Reference 10: The correct citation should be: “Electrooculogram Analysis and Development of a System for Defining Stages of Drowsiness Master's Thesis Project

in Biomedical Engineering, Linköping University, Dept. Biomedical Engineering, LiU-IMT-EX-351 Linköping 2003. Available: https://www.diva-portal.org/smash/get/diva2:673960/FULLTEXT01.pdf Test"

- Reference 19: The correct citation should be "Anxiety Detection Using Multimodal Physiological Sensing, 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece, 2021, pp. 1-4, doi: 10.1109/BHI50953.2021.9508589."
- Reference 23: Revising this citation is advised as searching on the internet shows error 404. The requested URL was not found on this server. Moreover, this is not a proper citation—give the edition number of the book (there are at least 5 editions) and publication year, as well as the page number of the cited data point about typical blink elapsed time.
- Reference 27: The correct citation should be "Hassanein, A.M.D.E., Mohamed, A.G.M.A. & Abdullah, M.A.H.M. Classifying blinking and winking EOG signals using statistical analysis and LSTM algorithm. Journal of Electrical Systems and Inf Technol 10, 44 (2023). https://doi.org/10.1186/s43067-023-00112-2."
- In general, citations need to be reviewed and added with consistency throughout the manuscript.

Acknowledgments

PREreview and JMIR Publications thank the authors 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

DS was a facilitator of this call and one of the organizers. No other competing interests were declared by the reviewers.

References

1. Dao J, Liu R, Solomon S, Solomon S. State anxiety biomarker discovery: electrooculography and electrodermal activity in stress monitoring. arXiv. Preprint posted online on Nov 26, 2024. [doi: 10.48550/arXiv.2411.17935]
2. Mitchell LA, MacDonald RAR, Brodie EE. Temperature and the cold pressor test. J Pain 2004 May;5(4):233-237. [doi: 10.1016/j.jpain.2004.03.004] [Medline: 15162346]

Abbreviations

EDA: electrodermal activity
EOG: electrooculography
s-anxiety: state anxiety
SHAP: Shapley Additive Explanations
STAI: State-Trait Anxiety Inventory

Edited by A Schwartz; submitted 03.02.25; this is a non–peer-reviewed article;accepted 03.02.25; published 03.03.25.

Please cite as:
Saderi D, Rasania S, Olatoye T, Savai SM, Mahmoud RSG, Medeiros V, Collier M
Peer Review of "State Anxiety Biomarker Discovery: Electrooculography and Electrodermal Activity in Stress Monitoring (Preprint)"
JMIRx Med 2025;6:e72093
URL: https://xmed.jmir.org/2025/1/e72093
doi:10.2196/72093

Peer Review of "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance"

Daniela Saderi[1]; Goktug Bender[2]; Toba Olatoye[3]; Arya Rahgozar[4], PhD; Uday Kumar Chalwadi[5]; Eudora Nwanaforo[6]; Paul Hassan Ilegbusi[7]; Sylvester Sakilay; Mitchell Collier

[1]PREreview, Portland, OR, United States
[2]McGill University, Montreal, ON, Canada
[3]University of Ilorin, Ilorin, Nigeria
[4]University of Ottawa, Ottawa, ON, Canada
[5]LSUHS, Shreveport, LA, United States
[6]Federal University of Technology, Owerri, Nigeria
[7]Ondo State College of Health Technology, Akure, Nigeria

Related Articles:

Companion article: https://www.medrxiv.org/content/10.1101/2024.08.09.24311777v1

Companion article: https://med.jmirx.org/2025/1/e73258

Companion article: https://med.jmirx.org/2025/1/e65263

(JMIRx Med 2025;6:e73264) doi:10.2196/73264

KEYWORDS

natural language processing; NLP; machine learning; ML; artificial intelligence; language model; large language model; LLM; generative pretrained transformer; GPT; pediatrics

This is the peer-review report for "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance."

This review is the result of a virtual collaborative live review organized and hosted by PREreview and JMIR Publications on October 25, 2024. The discussion was joined by 21 people: 2 facilitators, 1 member of the JMIR Publications team, and 18 live review participants, including 3 who agreed to be named here but did not contribute to writing this review: Nour Shaballout, Randa Salah Gomaa Mahmoud, and Samaila Jackson Yaga. The authors of this review have dedicated additional asynchronous time over the course of 2 weeks to help compose this final report using the notes from the live review. 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] seeks to determine how accurately and reliably a fine-tuned GPT-3 model can assist with differential diagnosis in pediatric cases within rural health care environments. Specifically, it examines whether the artificial intelligence (AI) model can match or approach the diagnostic accuracy of human physicians. By evaluating the model's diagnostic performance, the research aims to explore AI's potential to improve pediatric health care quality, reduce misdiagnosis, and support providers in underserved regions where accurate, timely diagnosis is critical for patient outcomes.

To address the research questions, the authors conducted a retrospective study using data from 500 pediatric cases from a multicenter rural pediatric health care organization in Central Louisiana, United States. The GPT-3 model was trained on 70% of the data, including symptoms and physician-provided differential diagnoses, and tested on the remaining 30%, achieving an accuracy of 87%, with sensitivity at 85% and specificity at 90%. These results were statistically comparable to human physicians, who had an accuracy of 91%. The findings suggest that AI can support clinical decision-making in pediatric care, especially in resource-constrained environments where access to specialists is limited.

The research addresses critical gaps in pediatric care by exploring AI's potential to support clinical decision-making, particularly in resource-limited settings. It presents this with methodological details that enhance reproducibility and offer insights into AI applications in health care. The authors' transparency about limitations reflects research integrity, establishing a strong base for future studies. Furthermore, the focus on integrating AI into clinical workflows shows an understanding of practical challenges and underscores opportunities for advancing health care delivery through

technology. However, the study presents some notable weaknesses, including a lack of assessment of patient outcomes and insufficient clarity in its methodology, indicating areas for future research and improvement. Below, we list specific concerns and recommendations on how to address them.

List of Major Concerns and Feedback

Concerns With Techniques and Analyses

- Model choice: It is unclear why a specific generative AI model (ie, GPT-3, DaVinci version) was chosen for this study. Was the GPT-3 model (DaVinci version) selected due to its extensive use in medical AI research, or was it chosen to facilitate comparison with previous studies? A statement explaining the choice of the AI model would significantly improve the reader's understanding of the study's context and its relationship to previous research.
- Normality test: The study does not address whether data normality was assessed before statistical analysis. Determining the distribution of the data is key to selecting the appropriate statistical test to analyze such data. The Kolmogorov-Smirnov test could aid in understanding data distribution, specifically testing for normality. If the data is not found to meet normality criteria, nonparametric methods should be applied. Including a data normality assessment and explaining the choice of a particular statistical test would significantly strengthen the reliability of the study.
- Evaluation metrics: The study primarily uses specificity and sensitivity for evaluating large language model–generated responses, which may not capture the full quality of the outputs. Incorporating natural language processing metrics such as Recall-Oriented Understudy for Gisting Evaluation (ROUGE) and bilingual evaluation understudy (BLEU) can help assess the quality of generated responses more comprehensively. ROUGE measures the correspondence between the automatically generated response versus that of the human and what was expected. There are also issues associated with large language model generations of responses such as hallucination and the lack of attribution. Please specify or comment on how those and other issues were measured.
- Power analysis assumptions: The assumptions underlying the power analysis are unclear, particularly regarding how specific diagnoses affect this analysis. It is advised to elaborate on the power analysis methodology, including the rationale behind sample size choices and their implications for diagnosis variability.
- Sample size and generalizability: The sample size of 500 encounters may not adequately represent the broader pediatric population, particularly in diverse settings. Furthermore, using data from a single health care organization limits the applicability of findings to other settings. These limitations should be discussed, particularly how the validity of the results might change when it is tested with data from other health care centers. If possible, authors should mention and cite studies that reported on this effect. Additionally, future studies should consider expanding the sample size through multicenter collaborations or including data from patients with more diverse demographics to validate results across different health care environments thereby enhancing generalizability.

Details for Reproducibility of the Study

- Software and tools documentation: The authors describe using both Python (with scikit-learn) and IBM SPSS Statistics, but it is unclear what the software's sources are. Specifying sources for Python and scikit-learn (eg, "Python 3.8 [Python Software Foundation, Delaware, USA]") and clarifying the respective roles of Python and SPSS in the analyses would enhance transparency and allow for the reproducibility of the study.
- Detailed group descriptions: The demographics, specifically age group cases, are underspecified, limiting the reader's understanding of the study sample. Adding a table or descriptive text detailing subgroup demographics, including age and case counts would improve the study's interpretability and allow readers to better contextualize findings.
- Cross-validation across organizations: The model's reproducibility across various health care settings is not demonstrated. Evidence shows models often underperform with data from different sources. Including cross-organization validation and clearly acknowledging this limitation in the Discussion by citing relevant studies would enhance robustness. Furthermore, addressing this limitation in future work could pave the way for broader adoption and application of the model.
- Data and model specifics for replicability: The study would benefit from more thorough descriptions of dataset characteristics, fine-tuning model parameters, and preprocessing methods. For validation, consider adding multicenter dataset details. Adding this information would enable other researchers to replicate and build upon the study's findings, thereby enhancing its scientific contribution.
- Diagnostic exclusion or inclusion clarification: The preprocessing section does not clarify if physician diagnostics were included or excluded, leading to potential confusion for readers and impacting reproducibility. It would be helpful to know whether physician diagnostics were included in training and why. Clarifying this aspect would help standardize study replication and improve the study's transparency.

Figures and Tables

- Figure 1 is mentioned but not included in the article, which affects comprehension of the study design and findings. Please include Figure 1 or provide an alternative reference to explain the content of the missing figure. Figures are helpful for readers to quickly grasp complex methodologies and findings.

Ethics

- Data privacy: It is unclear whether a private or public version of GPT-3 was used, and if the latter, this raises potential Health Insurance Portability and Accountability Act (HIPAA) concerns. As was already pointed out above,

it is recommended that the version of GPT-3 used is specified, with additional clarification regarding data privacy practices if a public model was used. The addition of HIPAA considerations will enhance readers' confidence in the study's privacy protocols.
- Discussion of diagnostic risk: The discussion would benefit from a deeper exploration of diagnostic risks associated with the use of AI in health care and clinical decision-making settings. One example is the potential of AI models to perpetuate and affirm existing human biases thereby further exacerbating health disparities (one relevant citation could be Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit Med.* Jun 14, 2023;6(1):113 [doi: 10.1038/s41746-023-00858-z] [Medline: 37311802]). The study also raises important social considerations, such as respecting human agency, particularly for vulnerable populations. Addressing parental concerns about deferring decision-making to AI is crucial, as is ensuring a socially attuned approach to building trust and understanding.
- Lack of clarity on potential implementation in rural health care settings: The study could be strengthened by detailing how the AI model might be implemented in rural health care settings, including the specific challenges involved. Key considerations include the need for sufficient infrastructure (eg, electricity, internet) and the necessity of training health care providers unfamiliar with AI tools. Additionally, discussing both the potential impact (eg, improved diagnostic efficiency) and limitations (eg, handling incomplete data or overreliance on AI) would provide a more comprehensive road map for deployment in rural environments.

List of Minor Concerns and Feedback

- Data distribution gaps: No comparison of racial identity distribution between training and testing sets. Please consider adding a table or section on these demographic comparisons to ensure representation across subgroups.
- Data description and context: It would be helpful to know more information regarding how physicians were selected and their specific roles in the study.
- Departmental affiliations: Authors' affiliations lack specific department details, which limits transparency. Include departmental affiliations for authors to increase transparency and traceability. Adding departmental affiliations will provide context on the authors' expertise and institutional support.
- Funding transparency: The funding statement does not clearly specify whether the study was internally or externally funded. Explicitly state funding details, clarifying internal/external sources as applicable. Clear funding information will enhance transparency and address potential conflicts of interest.
- Approval number: While an ethical approval statement is present, it lacks the approval number, which is critical for ethical transparency. Please include the ethics approval number/code to ensure proper documentation and strengthen the study's validity and trustworthiness.
- Inconsistent data collection dates between the abstract and data collection section (lines 19 and 82)
- Missing figure (line 104).
- Need for more descriptive statistics (mean, median, quartiles, SD).
- Data distribution: Lack of comparison for racial/Hispanic identity distribution between training and testing sets. There's insufficient detail on age subgroup distribution.
- Clarification needed: The authors need to provide a deeper discussion of the power analysis methodology.
- The authors assessed that the distribution of age, gender, and chief complaints was similar between the training and testing sets. Suggest this to be cited to Table 5.
- Table 1: The abbreviations in the formula column should be identified in the table legend as "(FN: False Negative; FP: False Positive; TN: True Negative; TP: True Positive) (m)+1."
- Please clarify why GPT-3.5 or GPT-4 (instead of GPT-3) was not used despite being available at the time of the study.
- Line 103 states physicians were instructed to generate differential diagnoses. I thought this was obtained retrospectively. Please clarify.
- Line 152: Table 4 should be corrected to Table 3.
- Line 154: Table 5 should be corrected to Table 4.
- Line 200: Typo "may limit the of the finding."

Concluding Remarks

We thank the authors 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

DS contributed to writing this review and was a facilitator of this call and one of the organizers. No other competing interests were declared by other reviewers who participated in discussing the preprint during the live review.

Reference

1. Mansoor M, Ibrahim AF, Grindem D, Baig A. Large language models for pediatric differential diagnoses in rural health care: multicenter retrospective cohort study comparing GPT-3 with pediatrician performance. JMIRx Med 2025;6:e65263. [doi: 10.2196/65263]

Abbreviations

AI: artificial intelligence
BLEU: bilingual evaluation understudy
HIPAA: Health Insurance Portability and Accountability Act
ROUGE: Recall-Oriented Understudy for Gisting Evaluation

Edited by A Schwartz; submitted 28.02.25; this is a non–peer-reviewed article;accepted 28.02.25; published 19.03.25.

Please cite as:
Saderi D, Bender G, Olatoye T, Rahgozar A, Chalwadi UK, Nwanaforo E, Ilegbusi PH, Sakilay S, Collier M
Peer Review of "Large Language Models for Pediatric Differential Diagnoses in Rural Health Care: Multicenter Retrospective Cohort Study Comparing GPT-3 With Pediatrician Performance"
JMIRx Med 2025;6:e73264
URL: https://xmed.jmir.org/2025/1/e73264
doi:10.2196/73264

Peer Review of "Safety and Efficacy of Chimeric Antigen Receptor T-Cell Therapy for Recurrent Glioblastoma: An Augmented Meta-Analysis of Phase 1 Clinical Trials (Preprint)"

Vanessa Fairhurst[1]; Randa Salah Gomaa Mahmoud[2]; Toba Olatoye[3]; Sylvester Sakilay

[1]PREreview, -, Portland, OR, United States

[2]Faculty of Human Medicine, Zagazig University, Zagazig, Egypt

[3]University of Ilorin, Ilorin, Nigeria

Related Article:

Companion article: https://www.medrxiv.org/content/10.1101/2024.10.23.24316015v1

(JMIRx Med 2025;6:e71293) doi:10.2196/71293

KEYWORDS

CAR T-cell therapy; cancer; glioblastoma; brain tumor; meta-analysis; chimeric antigen receptor

This is a peer-review report for the preprint "Safety and Efficacy of Chimeric Antigen Receptor T-cell Therapy for Recurrent Glioblastoma: An Augmented Meta-Analysis of Phase 1 Clinical Trials."

This review is the result of a virtual collaborative live review discussion organized and hosted by PREreview and JMIR Publications on Dec 12, 2024. The discussion was joined by 11 people: 3 facilitators, 1 member of the JMIR Publications team, and 7 live review participants including 3 who agreed to be named but did not assist in compiling the final review: Eudora Nwanaforo, Kelechi Elechi, and Murtala Haruna Bawa. The authors of this review have dedicated additional asynchronous time over the course of 2 weeks to help compose this final report using the notes from the live review. 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] was designed to address the limitations of previous studies and evaluate the safety and efficacy of chimeric antigen receptor (CAR) T-cell therapy for recurrent glioblastoma. The results of this study are predictive rather than confirmatory. CAR T-cell therapy for glioblastoma was not predicted to significantly improve survival or achieve substantial complete responses. Stable disease rates were modest, while disease progression was notable. Adverse events, especially CAR T-cell therapy–related encephalopathy, raise safety concerns. Overall survival was 6.49 months in patients receiving CAR T-cell therapy after augmented analysis, and only 80% of patients exhibited this outcome. It was not statistically different from the median overall survival observed in patients with recurrent glioblastoma undergoing standard treatment, thereby indicating that CAR T-cell therapy, in its current form, does not offer substantially improved survival compared to standard treatments. Further trials and refinements are needed to enhance CAR T-cell therapy's effectiveness and safety in glioblastoma treatment. An interesting fact is that a novel statistical technique (augmented meta-analyses) was used in this study. It was a combination of a cross-sectional (quantitative) and augmented meta-analysis (qualitative).

List of Major Concerns and Feedback

Methods

Augmented Meta-Analysis

- This section is limited in its description of the methodology used in the study. It would be helpful to include more information on the machine learning model or language model used to generate the extra cases.
- The title and aim specify that the study focuses on recurrent glioblastoma, but this specificity is not reflected in the inclusion criteria. It would be helpful to adjust the inclusion criteria to explicitly state that the study is targeting patients with recurrent glioblastoma. This will align the methodology with the aim as stated.
- The inclusion criteria do not specify that patients are in phase 1 clinical trials, where safety is a primary focus. Clearly state in the inclusion criteria that patients are part of phase 1 clinical trials. This will provide context for the study's focus on safety.
- There is no reference to the earlier use of augmented meta-analysis in cancer or medical research, nor is it explicitly stated if this is a new application. If augmented meta-analysis has been previously applied, cite relevant references. If this is its first application, explicitly state so and highlight its novelty.

Results

Literature Review and Risk of Bias Assessment Section

- It would be helpful to add the details of Figure 1 and Table 1 that explain the details of the cause of exclusion, the

results of the Newcastle Ottawa Scale, which study reached the high-quality level, etc.

Discussion

- It is important to add a comparison between the mean overall survival for patients with glioblastoma who underwent CAR T-cell therapy and the median overall survival observed in patients receiving the standard protocol for recurrent glioblastoma treatment to the Results section, as this comparison is mentioned in the first paragraph of the Discussion section.

Reproducibility of the Study

- The data presented in the study are beneficial for reproducibility except for the augmented meta-analysis, which is hindered by the lack of clear documentation on the large language model settings.
- The details of the augmented meta-analysis are not available. Provide access to the source code or methodological details for augmented meta-analysis, either as supplementary material or a public repository link. Transparency will strengthen the study's reproducibility.

List of Minor Concerns and Feedback

Concerns With Techniques/Analyses

- Abbreviations like "IL-13Ralpha-2," "EGFRvIII," "HER2," and "HephA2" are not identified in the Included Study Characteristics section. Expand the abbreviations and provide their full names (eg, "Interleukin-13 Receptor Subunit Alpha-2") when first mentioned. This ensures clarity for readers not familiar with the terms.
- The last line of the large language model statement on page 16 does not explain how augmented meta-analysis was applied. Elaborate on how augmented meta-analysis was applied, especially in terms of methodology and integration with the study data.

Figures and Tables

- The screening section in Figure 1 is missing a rectangle to indicate the exclusion of 300 records. Update it using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart to include a rectangle that details the 300 excluded records and ensures the causes of exclusion are clearly stated.
- The reasons for exclusion are not detailed in the PRISMA flowchart. Follow PRISMA guidelines to specify the causes of exclusion, such as duplicates, irrelevance, or incomplete data, within the flowchart.
- Comments following Figure 1 are not in line with its instructions. Restructure the comments to follow the instructions and present the details of the research study accordingly.

Additional Comments

- No reference is provided for the trim-and-fill method mentioned in the augmented meta-analysis of overall survival (page 10). Cite a relevant source, such as [2] or another appropriate reference.
- The Cochrane Handbook (Part 2, Chapter 9) should be referenced in the Statistical Analysis section and its numbered reference cited in the text.
- References in the third paragraph of the Introduction mix meta-analyses and clinical trials without clear distinction. Rearrange and clarify the references while ensuring that references to meta-analyses and clinical trials are grouped and contextualized appropriately to avoid confusion.
- Repetition of the sentence "Egger's test for publication bias could not be performed since the number of included studies in this outcome was less than ten" could be avoided by mentioning it once in the Methods section as the total number of the included studies is 8.
- In addition, the repetition of the sentence "The wide range of the 95% confidence interval was suggestive of data sparsity, so augmented meta-analysis was indicated before making conclusions" could be avoided by mentioning it once in the Augmented Meta-Analysis section of the Methods.

Acknowledgments

PREreview and JMIR Publications thank the authors 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

VF was a facilitator of this call and one of the organizers. No other competing interests were declared by the reviewers.

References

1. Azzam AY, Morsy MM, Azab MA, et al. Safety and efficacy of chimeric antigen receptor T-cell therapy for recurrent glioblastoma: an augmented meta-analysis of phase 1 clinical trials. medRxiv. Preprint posted online on Oct 24, 2024. [doi: 10.1101/2024.10.23.24316015]
2. Shi L, Lin L. The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. Medicine (Baltimore) 2019 Jun;98(23):e15987. [doi: 10.1097/MD.0000000000015987] [Medline: 31169736]

Abbreviations

CAR: chimeric antigen receptor
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Schwartz; submitted 14.01.25; this is a non–peer-reviewed article;accepted 14.01.25; published 24.01.25.

Please cite as:
Fairhurst V, Mahmoud RSG, Olatoye T, Sakilay S
Peer Review of "Safety and Efficacy of Chimeric Antigen Receptor T-Cell Therapy for Recurrent Glioblastoma: An Augmented Meta-Analysis of Phase 1 Clinical Trials (Preprint)"
JMIRx Med 2025;6:e71293
URL: https://xmed.jmir.org/2025/1/e71293
doi:10.2196/71293

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org