Contents

Peer-Review Reports

Peer Review of “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19” (e24453)
Ahmed Hamed. ..................................................................................................................... 3

Peer Review of “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19” (e24481)
Susan Howlett. ..................................................................................................................... 5

Peer Review of “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study” (e24747)
Andy Chang. ......................................................................................................................... 7

Peer Review of “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study” (e24765)
Eric Abbott. .......................................................................................................................... 9

Peer Review of “Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study” (e25572)
Ross Gore. ................................................................................................................................ 11

Authors’ Response to Peer Reviewss

Authors’ Response to Peer Reviews of “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19” (e24485)
Moshe Rogosnitzky, Esther Berkowitz, Alejandro Jadad. .............................................................. 13

Author Response to Peer Reviews of “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study” (e24739)
Eric Luellen. .......................................................................................................................... 15
Author’s Response to Peer Review of “Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study” (e25573)
Tara Fusillo. .......................................................................................................................... 17

Viewpoint

No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19 (e19583)
Moshe Rogosnitzky, Esther Berkowitz, Alejandro Jadad. .......................................................... 19

Original Papers

A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study (e23582)
Eric Luellen. .......................................................................................................................... 33

Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study (e22470)
Tara Fusillo. .......................................................................................................................... 45
Peer Review of “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19”

Ahmed Abdeen Hamed¹, PhD
School of Cybersecurity, Data Science and Computing, Norwich University, Northfield, VT, United States

Related Articles:
Companion article: https://preprints.jmir.org/preprint/19583
Companion article: https://med.jmirx.org/2020/1/e24485/
Companion article: https://med.jmirx.org/2020/1/e19583/

(JMIRx Med 2020;1(1):e24453) doi:10.2196/24453

KEYWORDS
COVID-19; drug repurposing

This is a peer review submitted for the paper “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19.”

Round 1 Review

General Comments
The title of the manuscript presents the urgent topic of drug repurposing in the context of COVID-19. Clearly, research in this area is much needed, and the esteemed authors have put it well: “no time to lose.” The manuscript, however, suffers from some significant and obvious issues that can be summarized as follows:

Specific Comments

Major Comments
1. The esteemed authors didn’t use the preferred template. Particularly, the abstract section, if the template is followed, should summarize the research background, objectives, methods, results, and conclusions. These sections are not clear from the current abstract.
2. The manuscript is missing significant sections such as the Methods section, which is the most exciting part of any paper. It is not clear how the esteemed authors have summarized the literature (methods, experiments, tools, development environment, etc). The only methods that were mentioned was in reference number 67, but the paper has entirely missed this very significant section (among many others).
3. The manuscript is also missing significant content that fails to demonstrate how the analysis was conducted and how the results are presented. The manuscript does not have a single figure and has only one table.
4. The esteemed authors have immediately presented their classes (which I believe to be the Conclusion section) after the Background section. Since it does not follow the JMIR template, this manuscript does not make it possible for the information in this paper to be accessible to the reader.
5. The esteemed authors have presented a report or an early stage whitepaper that can potentially be turned into a research paper. However, in the current stage, I don’t believe it qualifies as a publishable manuscript (not without adding a section on methods, experiments, etc, and following the recommended template). I highly recommend to the esteemed authors to search the JMIR archives for keywords such as literature, COVID, and drug repurposing and find some publications that they can potentially use as a guide to present their work.

Conflicts of Interest
None declared.

https://xmed.jmir.org/2020/1/e24453
Peer Review of “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19”

Susan Howlett†, BSc, MSc, PhD, PDF
Department of Pharmacology, Dalhousie University, Halifax, NS, Canada

Related Articles:
Companion article: https://preprints.jmir.org/preprint/19583
Companion article: https://med.jmir.org/2020/1/e24485/
Companion article: https://med.jmir.org/2020/1/e19583/

(JMIRx Med 2020;1(1):e24481) doi:10.2196/24481

KEYWORDS
COVID-19; drug repurposing

This is a peer review submitted for the paper “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19.”

Round 1 Review

General Comments
Drug repurposing offers the chance to quickly move potentially effective drugs forward to treat COVID-19. The authors note that drug candidates should ideally have a track record of safety, affordability, and accessibility. The authors use the literature to support the use of four classes of generic drugs that target the pathophysiology of COVID-19: (1) the histamine H2 receptor antagonists cimetidine and famotidine; (2) the antiplatelet agent/phosphodiesterase inhibitor dipyridamole; (3) the phosphodiesterase-5 inhibitor sildenafil; and (4) the cholesterol-lowering agents fenofibrate and bezafibrate. These drugs are affordable and supported by decades of safety data. They also argue that this approach can be used to identify other drugs to address this urgent clinical need. The work is provocative and addresses a topic of great importance. Although it is generally well written, there are areas where this manuscript could be revised to improve clarity and readability, as outlined below.

Specific Comments

Major Comments

Overall
The authors use many nonstandard abbreviations. Also, many of these are used only once or a few times. To improve readability, especially for this review article, please greatly reduce the number of abbreviations, especially nonstandard ones that are hardly used. Also, please make sure that all of your abbreviations are defined.

There are a few new papers in this rapidly evolving area that the authors may wish to include in this review. These are (1) cimetidine and famotidine: an ongoing RCT for famotidine (NCT04370262; https://pubmed.ncbi.nlm.nih.gov/32446698/); (2) dipyridamole: an observational study (NCT04424901); (3) sildenafil citrate: not yet recruiting (NCT04489446).

Abstract
The abstract is a bit vague about the nature of the drugs used and the pathophysiology targeted. It would be important to mention the drug classes as well as (or instead of) the specific drugs of interest in the abstract. In addition, it would be helpful if the authors could mention some of the critical underlying pathophysiological mechanisms that these drugs target with respect to COVID-19 (eg, anti-inflammatory, antiviral, cardioprotective, etc).

Background section
As for the abstract, please refer to the drug classes as well as the specific drugs of interest.

Pathophysiology of COVID-19 section
In this section, it is not clear whether the authors mean inpatients or all affected individuals. In Table 1, it is easy to see fever in 79%-98% of all affected individuals, but it is more difficult to understand the 59% with sepsis refers to individuals in the community. Is this at presentation? These numbers need a better context. It would also be helpful to have references in the table. Rearranging Table 1 would improve readability.

System, Clinical Finding, Prevalence, Reference
Respiratory Fever 79%-98% X
Cough 58%-79% X
etc…
Also, in this section, the authors refer to elevation of various plasma inflammatory biomarkers in COVID-19. Please list these, or the major ones. Alternatively, please clarify if they differ from what is listed in Table 1.

Potential Therapies Within the Current Pharmacopoeia Section

On page 4, in the sentence starting with “All three drugs…,” the authors argue the benefits of strategies that utilize the safest drugs with pleiotropic effects to treat COVID-19. This is an interesting and important point. However, this sentence is too long to be easily understood. Please revise and expand to clarify meaning.

Table 2: This table would be improved by the addition of specific references.

Dipyridamole section

Page 3, top paragraph: The sentence starting with “Within the 200-400 mg…” needs a reference.

Page 4, second paragraph: For the sentence starting with “Examples include…,” please discuss/explain prednisolone and explain what is meant by “widening of the therapeutic window of glucocorticoid activity.”

Sildenafil section

The authors mention the ongoing phase 3 trial of sildenafil (100 mg daily for 14 days in patients with COVID-19 and give the trial number [NCT04304313]). It would be helpful to include trial numbers in Table 2.

Conclusions:
The sentence beginning with “The efforts we make now to facilitate…” is really long and complex. It is hard to follow as written. Please clarify.

Minor Comments

Please bold headings for Sildenafil and Fenofibrate / Bezafibrate for consistency.

Please be consistent with SARS-CoV—the V is not capitalized in all cases.

Page numbers are not sequential.

Conflicts of Interest

None declared.
Peer Review of “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study”

Andy Chang¹, BS, MS
School of Professional Studies, Northwestern University, Chicago, IL, United States

Related Articles:
Companion article: https://preprints.jmir.org/preprint/23582
Companion article: https://med.jmirx.org/2020/1/e24739/
Companion article: https://med.jmirx.org/2020/1/e23582/

(JMIRx Med 2020;1(1):e24747)  doi:10.2196/24747

KEYWORDS
infectious disease; SARS-CoV-2; COVID-19; public health; immunity: vaccinations; therapeutics; stem-cell growth factor-beta

This is a peer review submitted for the paper “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study.”

Comments for Authors/Editors
Eric, very clear and concise paper. Very minor suggestions/clarifications/edits only on the abstract, none that I would consider to be substantive.

1. Stylistically, and I went back and forth on this multiple times as I did each of my reviews, the abstract almost reads like it was written by a different person. Your “main” paper flowed beautifully well in what I’ve come to expect from you: you set the stage, you provide the necessary background, you explain succinctly what you did, and then you go into discussions about the potential implications of the work. The abstract almost reads like separate bullets in a white paper, and I realize that you’re going up against wording or space constraints so in the big picture and with resource management, this isn’t one of those tasks or comments that rises to the top.

2. In the conclusion portion of the abstract, do you think it would add value to call out or hat tip the following: (a) any of the limitations that you highlighted (maybe the top one or two?) and (b) the tie-in to economic and socioeconomic factors that you mentioned in the conclusion of the main paper? Again, I recognize that you may be going up against wording or space constraints so in the big picture and with resource management, this isn’t one of those tasks or comments that rises to the top.

Conflicts of Interest
None declared.
Peer-Review Report

Peer Review of “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study”

Eric Abbott¹, BASc EE, PEng, PE, MBA, MSEE
Medical Health Informatics Graduate Program, Northwestern University, Chicago, IL, United States

Corresponding Author:
Eric Abbott, BASc EE, PEng, PE, MBA, MSEE
Medical Health Informatics Graduate Program
Northwestern University
339 East Chicago Avenue
Chicago, IL
United States
Phone: 1 3125036950
Email: abbottericb@gmail.com

Related Articles:
Companion article: https://preprints.jmir.org/preprint/23582
Companion article: https://med.jmirx.org/2020/1/e24739/
Companion article: https://med.jmirx.org/2020/1/e23582/

(JMIRx Med 2020;1(1):e24765) doi:10.2196/24765

KEYWORDS
infectious disease; SARS-CoV-2; COVID-19; public health; immunity: vaccinations; therapeutics; stem-cell growth factor-beta

This is a peer review submitted for the paper “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study.”

Specific Comments

General Comments

This paper is excellent and very timely given its importance as it relates to supporting forthcoming mass vaccinations to address COVID-19 and potential prioritization of such vaccinations based on the study findings.

Conflicts of Interest
None declared.

Major Comments
1. None
2. None
3. None

Minor Comments

https://med.jmirx.org/2020/1/e24765/
Peer Review of “Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study”

Ross Gore¹, MSc, PhD
Virginia Modeling, Analysis and Simulation Center, Old Dominion University, Norfolk, VA, United States

Related Articles:
Companion article: https://preprints.jmir.org/preprint/22470
Companion article: https://med.jmir.org/2020/1/e25573/
Companion article: https://med.jmir.org/2020/1/e22470/

(JMIRx Med 2020;1(1):e25572) doi:10.2196/25572

KEYWORDS
coronavirus; SARS-CoV-2; pandemic; socioeconomic status; predictive model; health care resource allocation

This is a peer review submitted for the paper “Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study.”

Round 1 Review

General Comments
This study [1] proposes a repeatable modeling process to identify regional population centers with COVID-19 vulnerability using linear regression. The work is validated using data from states with high population densities, with New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania showing the strongest predictive results.

Specific Comments

Major Comments

• The authors describe their stepwise linear regression process in the Methods section. However, I would like to see more detail (probably supplied as supplementary material) on the trendline and scatterplot analysis to identify those variables most correlated with COVID-19 mortalities. It seems a more comprehensive cross-correlational analysis would be appropriate here; perhaps the trendline and scatterplot analysis is sufficient, but it is impossible to tell without additional details and results being provided.

• The paper would be improved if it was grounded by more analytical background material with references. I think a review of related epidemiological models applying stepwise regression in a similar model would help readers buy into this approach and use/cite the authors’ model. This is not my area of expertise but some suggestions include:


  • Yu, HYR, Ho, SC, So, KFE and Lo, YL. The psychological burden experienced by Hong Kong midlife women during the SARS epidemic. Stress and Health. 2005;21(3):177-184. doi.org/10.1002/smi.1051

• I commend the authors for the validation of their model, but I believe a section describing the model limitations should be added. Specifically, I’d like to see a set of conditions describing what the model is unable to do, or under what conditions its uses are invalid. I assume this list would include:

  • The model is only valid within the range of the data that have been observed thus far.

  • The model would only need to be revalidated for each additional state not included in the Results section for which it is applied.
In addition, I think the authors should note that even though they checked for correlation between subcomponents of the CDC Social Vulnerability Index and COVID-19 mortalities, that other data sets at the FIPS level (ie, American Community Survey data and census data) might yield different results. In particular, there may be many variables correlated with one another that correlate with COVID-19 mortalities. GGore et al [2] showed that it can be very difficult to tease apart specific population demographic measures at a granular level (ie, city or FIPS) into a linear regression model since so many of these variables correlate highly with one another. Making this point and citing the paper would help establish the context under which the model was built and the conditions in which it is valid.

**Minor Comments**

- In the replication crisis error, an anonymized version of the data provided in the paper, the scripts used for analysis, and the scripts used to create the figures for the paper need to be provided to both the reviewers and to the readership. This ensures a completely transparent analysis and makes the authors’ paper significantly more impactful as other researchers can build off (and cite) the paper.
- Figures 2 and 3 use a contrast between red and green to convey meaning. As many as 8% of men and 0.5% of women are affected with the common form of red-green color blindness that creates no contrast between the two colors. The figure would be improved if a different color choice is used. There are numerous options that offer superior readability [3].

**Round 2 Review**

My comments have been addressed; the paper is now suitable for publication.

**Conflicts of Interest**

None declared.

**References**


©Ross Gore. Originally published in JMIRx Med (https://med.jmirx.org), 02.12.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://med.jmirx.org/, as well as this copyright and license information must be included.
Authors’ Response to Peer Reviews of “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19”

Moshe Rogosnitzky¹; Esther Berkowitz¹, MBChB, MA; Alejandro R Jadad², MD, DPhil, FRCPC, FCAHS, FRSA, LLD

¹Medinsight Research Institute, Rehovot, Israel
²Program in Impactful Giving, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Moshe Rogosnitzky
Medinsight Research Institute
Pekeris 4
Weizmann Science Park
Rehovot, 7670204
Israel
Phone: 972 522945520
Email: moshe@medinsight.org

Related Articles:
Companion article: https://preprints.jmir.org/preprint/19583
Companion article: https://med.jmirx.org/2020/1/e24453/
Companion article: https://med.jmirx.org/2020/1/e24481/
Companion article: https://med.jmirx.org/2020/1/e19583/

(JMIRx Med 2020;1(1):e24485) doi:10.2196/24485

KEYWORDS
COVID-19; drug repurposing

Author response to peer reviews for “No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19.”

Response to Round 1 Reviews

Response to Reviewers
The authors of the manuscript are grateful to the editor and reviewers for their invaluable input and feedback. We have taken the necessary steps to address the comments given and will be providing a summary of the changes made to the manuscript here. We also ensured that we revised the manuscript in line with common formatting/editorial guidelines as listed in JMIRx’s “Instructions for Authors” guide.

Reviewer E
The paper is a narrative review that consolidates the extensive evidence in the literature supporting the potential efficacy of six generic drugs (as discussed in the manuscript) in the management of COVID-19 patients. As such, it does not align with the “Introduction, Methods, Results, and Discussion-IMRD” structure that will be expected from an original article. Also, the current structure of the manuscript is in line with similar narrative reviews already published on JMIRx.

Reviewer F
Overall
All nonstandard abbreviations have been removed from the manuscript and use of abbreviations reduced significantly. Also, the authors have ensured that each abbreviation used is well defined at the point of first use.

We have included new papers (clinical trials) that have been generated since the manuscript was first submitted, as recommended.
Abstract
We have included the drug class for each drug and summarized the pathophysiological mechanism through which the drugs may help improve treatment outcomes in patients with COVID-19.

Background
Similar to the changes made to the abstract, we also explicitly mentioned the class to which each of the six drugs belongs as well as a summary of their underlying pathophysiological mechanism.

Pathophysiology of COVID-19 section
We have stated explicitly that the percentages reported are from hospitalized patients rather than individuals in the community.
We have also included relevant references in the table and changed the orientation of the table, as recommended.
As suggested, we have listed the major plasma inflammatory biomarkers that are abnormal in COVID-19, and these are the same as listed in Table 1.

Potential Therapies Within the Current Pharmacopeia section
We have restructured the sentence starting with “All three drugs...” to improve clarity and comprehension.
With respect to Table 2, we have added references from which the data presented for each drug have been sourced.

Dipyridamole section
We added a reference for the sentence starting with “Within the 200-400 mg...”
We have restructured the sentence starting with “Examples include...” first to improve the overall clarity of the sentence, and second to provide an explanation for “widening the therapeutic window of glucocorticoid activity.”

Sildenafil section
We have included trial numbers/identifiers in Table 2 for relevant ongoing clinical trials relating to sildenafil.

Conclusions
We modified the sentence beginning with “The efforts we make now to facilitate...” to improve its overall clarity.

Other minor comments
We made the headings for Sildenafil and Fenofibrate / Bezaafibrate bold for consistency with other similar headings.
The “V” in SARS-CoV-2 has been capitalized in all cases.
All page numbers are now sequential.
Authors' Response to Peer Reviews

Author Response to Peer Reviews of “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study”

Eric Luellen¹, MS, MPH
Bioinformatix, Interlochen, MI, United States

Corresponding Author:
Eric Luellen, MS, MPH
Bioinformatix
Faculty Lane, Box 3628
Interlochen, MI, 49643
United States
Phone: 1 4129157468
Email: eluellen@bioinformatix.io

Related Articles:
Companion article: https://preprints.jmir.org/preprint/23582
Companion article: https://med.jmirx.org/2020/1/e24747/
Companion article: https://med.jmirx.org/2020/1/e24765/
Companion article: https://med.jmirx.org/2020/1/e23582/

(JMIRx Med 2020;1(1):e24739)  doi:10.2196/24739

KEYWORDS
infectious disease; SARS-CoV-2; COVID-19; public health; immunity: vaccinations; therapeutics; stem-cell growth factor-beta

Author response to peer review reports for “A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study.”

Response to Round 1 Reviews

Regarding reviewer G’s feedback that the abstract read like bullets to a white paper and was stylistically different, and inferior, to the flow of the main paper, the abstract was revised to be more of a narrative while honoring the mandated structured subheadings. Moreover, the reviewer’s suggestion to include limitations and potential socioeconomic impacts of the results similarly improved the impact of the paper and contextualized its findings. The author is grateful for this insightful feedback because it helped improve the readability of the abstract, and may encourage more researchers, practitioners, and journalists to read the paper.

Regarding reviewer H’s feedback to be consistent with scientific nomenclature (SARS-CoV-2 or the more colloquial COVID-19), the paper was revised to note the alternative colloquial term once in the title and once in the text, and corrected all entries to the more medically and scientifically correct name, SARS-CoV-2. Again, the author is grateful for this constructive criticism because, in addition to consistency, it may impact readers’ inferences about the training of the author and scientific accuracy of the paper and results.
Authors’ Response to Peer Reviews

Author’s Response to Peer Review of “Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study”

Tara Fusillo
John F Kennedy High School, Bellmore, NY, United States

Corresponding Author:
Tara Fusillo
John F Kennedy High School
3000 Bellmore Ave
Bellmore, NY, 11710
United States
Phone: 1 516 992 1400
Email: ifusil@optonline.net

Related Articles:
Companion article: https://preprints.jmir.org/preprint/22470
Companion article: https://med.jmirx.org/2020/1/e25572/
Companion article: https://med.jmirx.org/2020/1/e22470/

(JMIRx Med 2020;1(1):e25573) doi:10.2196/25573

KEYWORDS
coronavirus; SARS-CoV-2; pandemic; socioeconomic status; predictive model; health care resource allocation

The author of the manuscript [1] is grateful to the editor and reviewers for their invaluable input and feedback.

Response to Round 1 Reviews

Specific Comments
Major Comments
• The author has added a correlation matrix to the supplemental materials.
• The author has added several examples of stepwise regression model use to the Introduction section to help ground readers in the validity of this approach in similar applications.
• The author has made the suggested points in the revised Limitations section of the Discussion.

Minor Comments
• All data sets and regression model details have been added for each state to the Multimedia Appendices section.
• The author has changed the conditional formatting to red and blue in Figure 3.

Reference
Edited by G Eysenbach; submitted 06.11.20; this is a non–peer-reviewed article; accepted 06.11.20; published 02.12.20.

Please cite as:
Fusillo T
Author’s Response to Peer Review of “Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study”
JMIRx Med 2020;1(1):e25573
URL: https://med.jmirx.org/2020/1/e25573/
doi:10.2196/25573
PMID:

©Tara Fusillo. Originally published in JMIRx Med (https://med.jmirx.org), 02.12.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://med.jmirx.org/, as well as this copyright and license information must be included.
No Time to Waste: Real-World Repurposing of Generic Drugs as a Multifaceted Strategy Against COVID-19

Moshe Rogosnitzky1; Esther Berkowitz1, MBChB, MA; Alejandro R Jadad2, MD, DPhil, FRCPC, FCAHS, FRSA, LLD

1Medinsight Research Institute, Rehovot, Israel
2Program in Impactful Giving, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Moshe Rogosnitzky
Medinsight Research Institute
Pekeris 4
Weizmann Science Park
Rehovot, 7670204
Israel
Phone: 972 522945520
Email: moshe@medinsight.org

Abstract

Real-world drug repurposing—the immediate “off-label” prescribing of drugs to address urgent clinical needs—is an indispensable strategy gaining rapid traction in the current COVID-19 crisis. Although off-label prescribing (ie, for a nonapproved indication) is legal in most countries, it tends to shift the burden of liability and cost to physicians and patients, respectively. Nevertheless, in urgent public health crises, it is often the only realistic source of a meaningful potential solution. To be considered for real-world repurposing, drug candidates should ideally have a track record of safety, affordability, and wide accessibility. Although thousands of such drugs are already available, the absence of a central repository of off-label uses presents a barrier to the immediate identification and selection of the safest, potentially useful interventions. Using the current COVID-19 pandemic as an example, we provide a glimpse at the extensive literature that supports the rationale behind six generic drugs, in four classes, all of which are affordable, supported by decades of safety data, and pleiotropically target the underlying pathophysiology that makes COVID-19 so dangerous. Having previously fast-tracked this paper to publication in summary form, we now expand on why cimetidine/famotidine (histamine type-2 receptor antagonists), dipyridamole (antiplatelet agent), fenofibrate/bezafibrate (cholesterol/triglyceride-lowering agents), and sildenafil (phosphodiesterase-5 inhibitor) are worth considering for patients with COVID-19 based on their antiviral, anti-inflammatory, renoprotective, cardioprotective, and anticoagulation properties. These examples also reveal the unlimited opportunity to future-proof public health by proactively mining, synthesizing, and cataloging the off-label treatment opportunities of thousands of safe, well-established, and affordable generic drugs.

(JMIRx Med 2020;1(1):e19583) doi:10.2196/19583

KEYWORDS
COVID-19; drug repurposing; fibrates; histamine type-2 receptor antagonists; cimetidine; famotidine; fenofibrate; bezafibrate; dipyridamole; sildenafil
**Introduction**

Since the first report of a viral pneumonia of unknown cause in Wuhan, China, in December 2019, followed by the identification of the virus SARS-CoV-2 and the designation of the disease it causes as COVID-19, we have witnessed the rapid development of a pandemic that has become a global public health crisis. Although reported mortality rates are between <1% and 27% depending on factors including age, gender, health status, and geographic location, this is likely an underestimation due to underreporting and limited serological testing [1]. With no approved preventive or therapeutic treatments available, the scale and human impact of the COVID-19 outbreak is daunting.

In this paper, we present candidates for a multifaceted approach to the management of COVID-19, based on repurposing tried and tested, affordable, widely available drugs with proven long-term safety, and mechanisms of action that address the underlying pathophysiology of the disease. Having recently published a short summary of our thesis [2], the current paper expands on why cimetidine or famotidine (histamine type-2 receptor antagonists), dipyridamole (antiplatelet agent), fenofibrate or bezafibrate (cholesterol/triglycerides-lowering agents), and sildenafil (phosphodiesterase-5 inhibitor) are worth considering for patients with COVID-19. The goal is to enable the rapid introduction of potentially beneficial, low-risk interventions. We also emphasize the urgency of redoubling efforts to mine, synthesize, and catalog the considerable existing body of evidence for promising treatments, in order to future-proof public health, based on robust science.

**Pathophysiology of COVID-19**

COVID-19 is characterized by prominent early respiratory signs and symptoms, including fever, cough, fatigue, and shortness of breath, that may deteriorate to acute respiratory distress syndrome (ARDS), coagulopathy, multiorgan failure, and other life-threatening sequelae (Table 1) [3,4]. On lung imaging, consolidation, ground glass opacity, and pulmonary infiltration are evident [3].

<table>
<thead>
<tr>
<th>Table 1. Common clinical findings, complications, and laboratory abnormalities in patients with laboratory-confirmed COVID-19. CRP: C-reactive protein.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respiratory [3,4]</strong></td>
</tr>
<tr>
<td>Fever</td>
</tr>
<tr>
<td>Cough</td>
</tr>
<tr>
<td>Respiratory failure</td>
</tr>
<tr>
<td>Acute respiratory distress syndrome</td>
</tr>
<tr>
<td>Sputum production</td>
</tr>
<tr>
<td><strong>Cardiovascular [3]</strong></td>
</tr>
<tr>
<td>Cardiac failure</td>
</tr>
<tr>
<td>Septic shock</td>
</tr>
<tr>
<td><strong>Multisystem [3,4]</strong></td>
</tr>
<tr>
<td>Sepsis</td>
</tr>
<tr>
<td>Fatigue</td>
</tr>
<tr>
<td>Coagulopathy</td>
</tr>
<tr>
<td><strong>Laboratory abnormalities [3,4]</strong></td>
</tr>
<tr>
<td>Lactate dehydrogenase &gt;245 U/L</td>
</tr>
<tr>
<td>Procalcitonin &lt;0.1 ng/mL</td>
</tr>
<tr>
<td>D-dimer &gt;1 μg/mL</td>
</tr>
<tr>
<td>Lymphopenia</td>
</tr>
<tr>
<td>Aspartate aminotransferase &gt; 40 U/L</td>
</tr>
<tr>
<td>Leucopenia</td>
</tr>
<tr>
<td>Raised CRP</td>
</tr>
</tbody>
</table>

*SARS-CoV-2 has also been isolated from feces, urine, blood, and ophthalmic secretions [5]; COVID-19 affects extrapulmonary organs and systems in ways that contribute significantly to overall morbidity and mortality. In a retrospective cohort study of 191 hospitalized patients with COVID-19 in Wuhan, sepsis was reported in 112 (59%) patients admitted to the hospital, while respiratory failure (54%), ARDS (31%), heart failure (23%), and septic shock (20%) were not available.*
reported in ≥20% of patients, significantly more frequently among those who died than in survivors (all P<.001) [3]. Similarly, coagulopathy (19%) and acute cardiac (17%) and renal (15%) injury were widely observed, with a significantly higher incidence in nonsurvivors [3]. The fact that underlying cardiovascular, pulmonary, and renal disease have been associated with significantly increased mortality in COVID-19 patients and that abnormalities of various plasma inflammatory biomarkers (e.g., lymphocyte count, C-reactive protein (CRP), procalcitonin, D-dimer, and aspartate aminotransferase) appear to be widespread [3,4,6,7] highlights the multisystem nature of the disease and suggests that immune-mediated cytokine signaling and development of cytokine storm play a key role in driving disease progression [7].

Laboratory findings support the diverse effects of SARS-CoV-2, demonstrating, among other derangements, that leukopenia, lymphopenia, and thrombocytopenia, as well as elevated lactate dehydrogenase, CRP, and D-dimer, are significantly associated with a more severe course of disease [3,4,6,8]. Viral load also correlates strongly with disease severity (lung injury in particular) [9], while virus-induced endothelial dysfunction contributes to acute cardiac events that are a recognized complication of COVID-19 [10]. Although little attention has focused specifically on disturbed coagulation, evidence suggests that COVID-19 leads to profoundly altered coagulation function, with raised D-dimer, fibrinogen, and fibrin/fibrinogen degradation products [11,12].

Table 2. Approved indications and recognized physiological effects of drugs that could be considered for repurposing in patients with COVID-19.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Current indications</th>
<th>Proposed dose</th>
<th>Demonstrated effects</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cimetidine or famotidine [23-46]</td>
<td>Symptomatic management of GERDb</td>
<td>Cimetidine 200 mg four times daily</td>
<td>✓✓✓✓✓✓</td>
<td>Establish baseline prolactin levels and monitor periodically.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Famotidine 20–40 mg twice daily</td>
<td>✓✓✓✓✓</td>
<td>May increase serum concentrations of other drugs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓✓✓✓✓✓✓</td>
<td>Reduces dipyridamole absorption.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓✓✓✓✓✓✓</td>
<td>Relevant trials: NCT04504240 and NCT04370262.</td>
</tr>
<tr>
<td>Dipryidamole [47-79]</td>
<td>Antithrombotic following cardiac valve replacement</td>
<td>75 mg thrice daily OR 50–100 mg once weekly</td>
<td>✓✓✓✓✓✓</td>
<td>May cause headaches during the first week of use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓✓✓✓✓✓✓</td>
<td>Taking with foods or antacids halves absorption.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓✓✓✓✓✓✓</td>
<td>Relevant trials: NCT04391179, NCT04424901, and NCT04410328.</td>
</tr>
<tr>
<td>Fenofibrate or bezafibrate [80-97]</td>
<td>Dyslipidemia</td>
<td>Fenofibrate ≤200 mg/day</td>
<td>✓✓✓✓✓✓</td>
<td>Significant reduction in D-dimer and fibrinogen usually seen in days.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bezafibrate 400 mg daily</td>
<td>✓✓✓✓✓</td>
<td>Relevant trial: NCT04517396.</td>
</tr>
<tr>
<td>Sildenafil citrate [98-112]</td>
<td>Erectile dysfunction</td>
<td>25 mg twice daily, on an empty stomach</td>
<td>✓✓✓✓✓</td>
<td>Avoid grapefruit juice (increases sildenafil levels).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓✓✓✓✓</td>
<td>Cimetidine/famotidine increases sildenafil concentration. If combined, consider lower sildenafil dose —even 12.5 mg twice daily.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓✓✓✓✓</td>
<td>Relevant trials: NCT04304313 and NCT04489446.</td>
</tr>
</tbody>
</table>

*Demonstrated effects are: A - preliminary efficacy in COVID-19 patients; B - anti-inflammatory effect; C - antiviral effect; D - anticoagulant effect; E - cardioprotective effect; F - renoprotective effect; G - lung protective effect.

bGERD: gastroesophageal reflux disease.

**Potential Therapies Within the Current Pharmacopoeia**

Widespread attention has been given to repurposing antimalarial chloroquine/hydroxychloroquine, the antibiotic azithromycin and, most recently, the antiparasitic agent ivermectin, for targeting COVID-19. All three drugs possess both anti-infective and immune-modulating properties [13-19]. While clinical evidence supporting their use individually or in combination in patients with COVID-19 have so far been inconclusive [13-19], partly due to methodological limitations, or are yet unavailable [20-22], this multitherapeutic approach of utilizing the safest drugs with pleiotropic effects is essential to reduce morbidity and mortality arising from COVID-19 infection. Indeed, a large number of approved drugs have mechanisms of action that could be harnessed to address the pathophysiology of COVID-19. Ideal choices would be safe and widely available generic drugs that are affordable in any setting, and especially for under-resourced populations. Below we summarize the safety profiles and rationale for repurposing several generic drugs that have demonstrated antiviral, anti-inflammatory, and/or cardio-, lung- or renal-protective effects. Some also lower elevated fibrinogen and D-dimer, which are associated with hypercoagulability and may contribute to the multiorgan failure seen in patients with COVID-19. Table 2 summarizes the physiological effects of these agents as they relate to potential benefits in patients with COVID-19.
Cimetidine and Famotidine

The histamine type-2 receptor antagonists (H₂RAs) cimetidine and famotidine were approved by the US Food and Drug Administration (FDA) in 1977 and 1986, respectively, and both have been widely used, for decades, for prevention and symptomatic management of gastroesophageal reflux disease (GERD) [23]. Ranitidine, another commonly used H₂RA, will soon be largely unavailable in the United States following an FDA recall based on high levels of a contaminant. Cimetidine is approved at daily doses of 200-400 mg for heartburn relief, and up to 1600 mg for the short-term treatment of erosive GERD, while famotidine is approved at a dose of 10-20 mg twice daily (bid) for GERD, and 20-40 mg bid for erosive esophagitis [23]. Both drugs are available over the counter in the United States and much of the world, and are generally well tolerated, with most adverse events reported in <1% of patients. Drug-drug interactions that may delay metabolism of other agents due to interaction with the cytochrome P450 system, limited antidote effects, and stimulation of prolactin are recognized in particular with cimetidine. Prolactin levels should therefore be established at baseline and monitored periodically.

Antiviral and Immunomodulatory Activity

Beyond their role as gastric acid reducers, H₂RAs have powerful modulatory effects on innate and adaptive immunity by interfering with the effects of histamine on a range of leukocytes. As such, they reverse histamine-mediated immunosuppression by stimulating the effector functions of a wide range of T and B cells [24]. The resulting antiviral effects have been demonstrated in small studies in patients with herpes simplex virus (HSV) [25] and herpes zoster infection [26], and with human papillomavirus–related disorders [27]. In preclinical studies, cimetidine boosted immune cellular response when used as an adjuvant for viral vaccines for hepatitis B virus [28-31] and suppressed HIV replication in vitro [32]. Furthermore, intravenous ranitidine significantly increased the antibody response to vaccination in patients receiving tetanus toxoid before major abdominal surgery [33] and in patients with B-cell chronic lymphocytic leukemia receiving tetanus toxoid–conjugated Haemophilus influenzae type b vaccine [34,35]. Beyond antiviral effects, H₂RAs have shown immunomodulatory effects in a range of cancers and allergic diseases, bone resorption, and during recovery from burn injury [24].

Cardiovascular Protective Effects

H₂RAs also demonstrate a number of cardioprotective effects. A meta-analysis of 10 randomized controlled trials in patients with congestive heart failure, most of whom used famotidine, showed that orally administered H₂RAs were associated with significant negative inotropic and chronotropic effects (reduction in heart rate vs placebo; P=0.02), and also significantly decreased blood pressure and increased cardiac efficiency, presumably reducing myocardial oxygen requirement [36]. In another study, in critically ill patients, a single intravenous infusion of cimetidine, 200 mg, reduced systolic, diastolic, and mean arterial blood pressure, and raised heart rate [37]. High-dose intravenous cimetidine (200 mg four times daily [qid]) administered after elective cardiac bypass surgery was also shown to reduce levels of proinflammatory mediators (neutrophil elastase, interleukin-8, CRP) with no adverse effects, suggesting the potential to improve cardiac outcomes under certain physiological conditions [38]. Furthermore, H₂RAs also strongly inhibit platelet aggregation in vitro [39,40] and ranitidine, in combination with hydrocortisone, has been shown to reduce complications after arterial thrombolyis in pediatric patients who developed arterial obstruction after cardiac catheterization [41]. These agents may therefore exert stabilizing effects on coagulation in patients with disturbed clotting function, the caveat being the potential for thrombocytopenia and/or hemolytic anemia [42-46].

Dipyridamole

The antiplatelet agent and phosphodiesterase inhibitor dipyridamole was first approved in 1961 and is indicated in the United States at doses of 75-100 mg qid with warfarin to decrease thrombotic risk following cardiac valve replacement [47]. It is also sold in the United States as a combined product, Aggrenox (aspirin 25 mg/extended-release dipyridamole 200 mg), which was approved in 1999 and is taken twice daily to reduce stroke risk [48]. Outside the United States, including in Europe, dipyridamole is available as a single agent; in Russia it is also approved as an antiviral agent [49]. Within the 200-400 mg daily dose range, dipyridamole is considered safe based on decades of clinical experience: adverse events are usually limited and transient, the most common being dizziness, gastrointestinal disturbance, headache, and skin rash [50]. The pleiotropic effects of dipyridamole derive from increased intracellular levels of cyclic adenosine monophosphate (cAMP) and cyclic guanine monophosphate (cGMP), which lead to anti-inflammatory, antioxidant, anticoagulation, and vasodilatory effects [51].

Antiviral Effects

Dipyridamole possesses very broad spectrum antiviral activity as shown in numerous preclinical studies that demonstrated efficacy, alone or as a potentiator of other agents, against HSV, HIV, variella zoster, cytomegalovirus, and mengovirus, as well as a range of viruses from the picornavirus, togavirus, orthomyxovirus, paramyxovirus, and pox virus families [52-56]. Induction of interferon responses have been identified as an important contributory factor to its antiviral effects [57]. A number of clinical studies have confirmed the antiviral effects of dipyridamole which, at a dose of 8-100 mg weekly, significantly reduced the risk of acute respiratory diseases, including influenza, when administered prophylactically to at-risk individuals [57-59].

Effects in Patients With COVID-19

A recent study in patients with COVID-19 in China illustrates that dipyridamole suppressed SARS-CoV-2 replication in vitro, induced potent antiviral immunity, and improved survival in a mouse model of pneumonia [60]. In a clinical study of 12 COVID-19 patients that was conducted alongside these preclinical investigations and reported within the same publication, dipyridamole increased lymphocyte and platelet counts, decreased D-dimer levels, and markedly improved clinical outcomes when dosed at 50 mg three times daily for 1
in plasma and may lead to increased drug metabolism and decreased therapeutic effects. Therefore, caution is advised when initiating or adjusting the dose of fenofibrate in patients who are already taking bemafenac. A dose adjustment might be necessary to maintain therapeutic levels and avoid potential drug interactions. 

Clinical Studies of Fenofibrate

Numerous clinical studies have been conducted to evaluate the efficacy and safety of fenofibrate in various patient populations. One study reported on the effectiveness of fenofibrate in reducing triglycerides and improving insulin sensitivity in patients with impaired fasting glucose or type 2 diabetes [83]. Another study demonstrated the beneficial effects of fenofibrate on cardiovascular outcomes in patients with dyslipidemia and type 2 diabetes [84]. These studies provide evidence for the role of fenofibrate in reducing cardiovascular risk and improving insulin sensitivity in high-risk populations.

Conclusion

Fenofibrate is a fibrate medication that has been widely used for the treatment of dyslipidemia. Its potential in reducing cardiovascular risk and its antihyperglycemic effects make it an attractive option for patients with multiple metabolic disorders. Further research is needed to explore its long-term safety and efficacy, as well as its role in other diseases such as non-alcoholic fatty liver disease and chronic kidney disease.
inflammation and apoptosis in human glomerular microvascular cells and reducing retinal microvascular inflammation [83,84], while bezafibrate decreased the number of circulating proinflammatory monocytes in patients with type 2 diabetes [85]. A beneficial modulatory role for fenofibrate in renal fibrosis and inflammation has also been proposed, with further studies required to elucidate the mechanisms involved [81].

Cardioprotective Effects

Preclinical studies provide evidence that fenofibrate can protect against cardiac ischemia-reperfusion injury and subsequent arrhythmias and heart failure, autoimmune myocarditis, and hypertension [81,86,87]. In the clinical setting, a meta-analysis of studies in which fibrates were used for primary prevention of atherosclerotic cardiovascular disease reported a significant 16% decrease in the combined outcome of cardiovascular death, nonfatal myocardial infarction, and nonfatal stroke [88] while a 12% reduction in these outcomes was reported when fibrates were used in secondary prevention [89].

Anticoagulation Effects

Robust data indicate that fibrates lower plasma fibrinogen levels to a significant degree, independent of their lipid-lowering effects [90-95]. In a meta-analysis of 22 trials, representing >2700 participants, fibrates (fenofibrate and bezafibrate) demonstrated a significantly greater effect than statins in lowering plasma fibrinogen concentrations (weighted mean difference –40.7 mg/dL, P < .001) [93]. Data from smaller studies focusing on the antiphylipidemic effects of fibrates (fenofibrate 200 mg daily; bezafibrate 400 mg daily) show concurrent, significant reductions in plasma fibrinogen levels [90-92,94]. Of note is a study that reported an increase in fibrinogen levels among patients taking statins [94], which could be a concern in elderly patients with COVID-19, many of whom are likely to be taking statins.

In patients with metabolic syndrome, which represents a hypercoagulable state accompanied by inflammation and endothelial dysfunction, fenofibrate reduced concentrations of thrombin-activatable fibrinolysis inhibitor, improved endothelial function [96], and significantly reduced fibrinogen and D-dimer concentrations [97], suggesting potential anticoagulant and cardiovascular protective effects. This potential was borne out in a short-term randomized controlled trial of patients with acute ST-elevation myocardial infarction, in whom bezafibrate lowered fibrinogen concentration more effectively than conventional therapy (P < .001), with significantly greater reductions in the incidence of angina (56% vs 4%, P < .001) and left ventricular failure (24% vs 4%, P = .049) [95].

Sildenafil Citrate

The phosphodiesterase-5 inhibitor (PDE5 inhibitor) sildenafil citrate is a vasodilator that was approved by the FDA in 1998 for the treatment of erectile dysfunction, at a dose of 25-100 mg once daily [98]. More recently, an indication for pulmonary arterial hypertension (PAH) was added in 2005, with an oral dosing of 5 or 20 mg thrice daily, or 2.5 mg or 10 mg as intravenous bolus [99]. In erectile dysfunction, orally administered sildenafil has an onset of action within 30 minutes, maximum effect at 1 hour, and duration of effect of 4-6 hours; in PAH, the pharmacodynamics are similar although peak effect occurs 1-2 hours after dosing, and blood pressure levels return to baseline levels within 8 hours [98,99]. The most common dose-dependent adverse reactions (≥5%) include headache, flushing, dyspepsia, visual disturbance, and nasal congestion. CYP3A4 inhibitors are known to potentiate sildenafil, while sildenafil potentiates the hypotensive effects of nitrates and alpha-blockers.

Sildenafil inhibits breakdown of cGMP through competitive binding at the phosphodiesterase binding site. It therefore influences platelet activation, proliferation of T cells, and production of proinflammatory cytokines, leading to a broad range of anti-inflammatory, antioxidant, vasodilatory, and other actions in many body systems [100]. An ongoing phase 3 trial of sildenafil, 100 mg daily for 14 days, in patients with COVID-19 (NCT04304313) will help clarify its potential benefits in this disease [101].

Immunomodulatory Effects

In vitro human studies indicate that sildenafil potentiates the ability of regulatory T cells to downregulate T effector cell proliferation, while clinical findings include reduced lymphocyte count and induction of malignant cell apoptosis in a patient with B-cell chronic lymphocytic leukemia and in patients with Waldenstrom’s macroglobulinemia [100]. It was hypothesized that these effects were mediated by synthesis and release of cytokines. Sustained increase in NO production, and decreased vascular inflammatory markers, have also been reported in patients with type 2 diabetes receiving sildenafil [102,103].

Cardiovascular Protective Effects

One-time and long-term administration of PDE5 inhibitors in patients at high cardiovascular risk can improve endothelial function, reduce inflammatory mediators, and increase endothelial regenerative capacity, which may be sustained for several months following treatment discontinuation, with potential applications in a range of cardiovascular disorders [104,105]. Cardioprotective effects include improved symptoms and cardiac contractility in patients with systolic heart failure, reduced myocardial infarct size, reduced blood pressure, and limitation of ischemia-driven ventricular arrhythmias, with reduction in cardiovascular events and mortality in high-risk patients [106-108]. In a British study that followed nearly 6000 men with type 2 diabetes over 7.5 years, the use of PDE5 inhibitors was associated with lower mortality risk overall (adjusted hazard ratio 0.54, P=.002) and in those with a history of acute myocardial infarction (heart rate=0.60, P=.001) [108]. These effects are believed to result from improved pulmonary circulation, as well as direct action on the myocardium, independent of the vasculature [106].

Lung-Protective Effects

Studies demonstrating sildenafil’s efficacy and tolerability in PAH continue to accrue, and a 2019 Cochrane systematic review and meta-analysis comprising 36 studies of nearly 3000 patients concluded that those with PAH who received PDE5 inhibitors were 22% less likely to die in the short-term than those receiving placebo [109]. Additionally, a network meta-analysis reported moderate-level evidence that sildenafil may reduce mortality
in idiopathic pulmonary fibrosis, an interstitial lung disease with high mortality [110]. A single case report of a 55-year-old physician with an atypical respiratory infection and apparently normal pulmonary arterial blood pressure who experienced marked symptomatic and functional improvement within 24 hours of starting tadalafil highlights the potential benefits of PDE5 inhibitors in this indication [111].

Renoprotective Effects

Preliminary evidence suggests that the clinical efficacy of PDE5 inhibitors in CKD extends beyond antihypertensive effects to active renoprotection. In preclinical studies, PDE5 inhibitors suppressed mesangial cell proliferation and extracellular matrix expansion, reduced renal cell apoptosis, and decreased oxidative stress and inflammation [107]. A post hoc examination of this class of medications in a randomized controlled trial also revealed improved kidney function and functional capacity, and a trend toward reduced mortality, in patients with PAH who received sildenafil treatment [112]. Few, if any, clinical studies of PDE5 inhibitors in patients with acute kidney disease have been published. Ongoing clinical trials (eg, NCT04304313) will shed further light on this and may reveal information that could be applied in the treatment of patients with COVID-19 [101).

Conclusions

Under the extraordinary COVID-19 pandemic conditions that have brought the world to the brink of an irreversible crisis, time is of the essence for the success of life-saving efforts. Until a vaccine is developed to treat this disease, the urgency of finding safe and effective treatments cannot be overstated. To ensure that patients with COVID-19 have rapid access to safe treatments, and to ensure the responsible use of available resources, it would be wise to mine the existing pharmacopeia for safe generic drugs that address the pathophysiologies underlying COVID-19. Moreover, beyond the current emergency there remains the likelihood of future re-emergence of another coronavirus or similar virus. The efforts we make now to facilitate access to information on the off-label applications of well-understood drugs, regardless of the manner in which the information has been discovered, are an investment in our future health that also addresses current needs. While clinical trials to assess efficacy will be important in due course, judicious use of one or more of these approved drugs, with caution toward potential interactions with concomitant medications, represents a rational and ethical approach that may prove effective in the short term. There is no time to waste and little to lose.

Epilogue

Since the initial submission of this article as a preprint in April 2020, new developments and evidence have emerged that further support the therapeutic potential of the drugs proposed in this paper for use in the treatment of COVID-19. The new developments and evidence are summarized below and are current as of August 31, 2020.

Dipyridamole

Research aimed at assessing the therapeutic potential of dipyridamole continues. One ongoing phase 3 clinical trial (ClinicalTrials.gov ID NCT04410328) randomized patients (n=132) to receive dipyridamole ER 200 mg and aspirin 25 mg orally/enterally plus standard care or standard care alone [113]. Researchers are also evaluating dipyridamole in two other ongoing clinical trials with a focus on determining the extent to which the drug can reduce excessive coagulation [114] and treat respiratory tract infection and circulatory dysfunction caused by SARS-CoV-2 [115] in hospitalized COVID-19 patients.

Famotidine

The therapeutic potential of famotidine (combined with cetirizine) in COVID-19 treatment was recently boosted by an American cohort study evaluating the efficacy of dual-histamine blockade in patients with COVID-19. In the study, hospitalized COVID-19 patients with severe-to-critical symptoms were treated with cetirizine 10 mg and famotidine 20 mg bid in addition to standard care. This combination reduced symptom progression when compared to published reports of COVID-19 patients [116]. The safety and efficacy of famotidine in COVID-19 is further supported by a case series of 10 US patients with COVID-19 who self-administered high-dose oral famotidine (80 mg thrice daily was the most frequent regimen used) for 11 days. All patients reported marked improvements in COVID-19–related symptoms, suggesting that high-dose oral famotidine is well tolerated and associated with improved patient-reported outcomes in nonhospitalized patients with COVID-19 [117].

Another case series of 14 COVID-19 hospitalized patients from Beloit Memorial Hospital, United States, reported improvement in supplemental oxygen requirements, ground-glass computed tomography findings, and serum levels of lactate dehydrogenase, ferritin, CRP, D-dimer, and lymphocytes in patients who received famotidine 80 mg qid plus celecoxib (as adjuvant therapy) [118]. This treatment combination was associated with a 100% survival rate. Similar clinical improvements have been reported by Freedberg et al [119] among hospitalized COVID-19 patients. Despite clinical evidence suggesting that famotidine may mitigate COVID-19, its mechanism of action remains a matter of debate. A recent study by Malone et al [120] suggests that the drug’s therapeutic action in COVID-19 involves on-target histamine receptor-H2 activity, which has face validity since the development of clinical symptoms involves dysfunctional mast cell activation and histamine release.

Fenofibrate

Researchers from the Hebrew University of Jerusalem, Israel, and Icahn School of Medicine at Mount Sinai (United States) studied the metabolic changes induced by SARS-CoV-2 infection in bronchial epithelial cells using lung biopsy samples from patients with COVID-19. The researchers reported the significant metabolic response in SARS-CoV-2–infected lungs in addition to changes in lipid metabolism and the induction of inositol-requiring enzyme-1 and RNA-activated protein kinase pathways of endoplasmic stress. The study showed that...
fenofibrate reversed the metabolic changes induced by SARS-CoV-2, blocking viral production and suppressing the pathogenesis of COVID-19 in lung tissue [121].

Sildenafil

A recent systematic review carried out by researchers from the University of Rome, Italy, consolidated evidence of the involvement of the NO-cGMP-PDE5 axis in the pathophysiology of COVID-19, presenting ongoing clinical trials aimed at modulating this axis, including the DEDALO (silDEnafil administration in DiAbetic and dysmetaboLic patients with COVID-19) trial [122]. Reviewed data indicate that PDE5 inhibitors could be effective in managing patients with COVID-19 by counteracting the Ang-II–mediated downregulation of the AT-1 receptor, exhibit action on monocyte switching, reducing proinflammatory cytokines and interstitial infiltration; and inhibit the transition of endothelial and smooth muscle cells to mesenchymal cells in the pulmonary artery, preventing clotting and thrombotic complications. With sildenafil’s low cost, well-established safety, wide availability, and efficacy arising from observational studies and clinical trials (including the new “Sildenafil in COVID-19” trial; ClinicalTrials.gov ID NCT04489446), it, and other PDE5 inhibitors, could potentially become key COVID-19 treatment options [122,123].

Conflicts of Interest

None declared.

References


98. Sildenafl citrate [Viagra]: Highlights of Prescribing Information. Pfizer, Inc. 2017. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/20895s039s042lbl.pdf [accessed 2020-04-10]


Abbreviations

ARDS: acute respiratory distress syndrome  
bid: twice daily  
cAMP: cyclic adenosine monophosphate  
cGMP: cyclic guanine monophosphate  
CKD: chronic kidney disease  
CRP: C-reactive protein  
CTAD: citrate-theophylline-adenosine-dipyridamole  
DEDALO: sildEnafil administration in DiAbetic and dysmetaboLic patients with COVID-19  
FDA: Food and Drug Administration  
GERD: gastroesophageal reflux disease  
H₂RA: histamine type-2 receptor antagonist  
HSV: herpes simplex virus  
NO: nitric oxide  
PAH: pulmonary arterial hypertension  
PDE: phosphodiesterase-5  
qid: four times daily
A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study

Eric Luellen¹, MS, MPH
Bioinformatix, Interlochen, MI, United States

Corresponding Author:
Eric Luellen, MS, MPH
Bioinformatix
Faculty Lane, Box 3628
Interlochen, MI, 49643
United States
Phone: 1 4129157468
Email: eluellen@bioinformatix.io

Related Articles:
Companion article: https://preprints.jmir.org/preprint/23582
Companion article: https://med.jmirx.org/2020/1/e24747/
Companion article: https://med.jmirx.org/2020/1/e24765/
Companion article: https://med.jmirx.org/2020/1/e24739/

Abstract

Background: Approximately 80% of those infected with COVID-19 are immune. They are asymptomatic unknown carriers who can still infect those with whom they come into contact. Understanding what makes them immune could inform public health policies as to who needs to be protected and why, and possibly lead to a novel treatment for those who cannot, or will not, be vaccinated once a vaccine is available.

Objective: The primary objectives of this study were to learn if machine learning could identify patterns in the pathogen-host immune relationship that differentiate or predict COVID-19 symptom immunity and, if so, which ones and at what levels. The secondary objective was to learn if machine learning could take such differentiators to build a model that could predict COVID-19 immunity with clinical accuracy. The tertiary purpose was to learn about the relevance of other immune factors.

Methods: This was a comparative effectiveness research study on 53 common immunological factors using machine learning on clinical data from 74 similarly grouped Chinese COVID-19–positive patients, 37 of whom were symptomatic and 37 asymptomatic. The setting was a single-center primary care hospital in the Wanzhou District of China. Immunological factors were measured in patients who were diagnosed as SARS-CoV-2 positive by reverse transcriptase-polymerase chain reaction (RT-PCR) in the 14 days before observations were recorded. The median age of the 37 asymptomatic patients was 41 years (range 8-75 years); 22 were female, 15 were male. For comparison, 37 RT-PCR test–positive patients were selected and matched to the asymptomatic group by age, comorbidities, and sex. Machine learning models were trained and compared to understand the pathogen-immune relationship and predict who was immune to COVID-19 and why, using the statistical programming language R.

Results: When stem cell growth factor-beta (SCGF-β) was included in the machine learning analysis, a decision tree and extreme gradient boosting algorithms classified and predicted COVID-19 symptom immunity with 100% accuracy. When SCGF-β was excluded, a random-forest algorithm classified and predicted asymptomatic and symptomatic cases of COVID-19 with 94.8% AUROC (area under the receiver operating characteristic) curve accuracy (95% CI 90.17%-100%). In total, 34 common immune
factors have statistically significant associations with COVID-19 symptoms (all c<.05), and 19 immune factors appear to have no statistically significant association.

**Conclusions:** The primary outcome was that asymptomatic patients with COVID-19 could be identified by three distinct immunological factors and levels: SCGF-β (>127,637), interleukin-16 (IL-16) (>45), and macrophage colony-stimulating factor (M-CSF) (>57). The secondary study outcome was the suggestion that stem-cell therapy with SCGF-β may be a novel treatment for COVID-19. Individuals with an SCGF-β level >127,637, or an IL-16 level >45 and an M-CSF level >57, appear to be predictively immune to COVID-19 100% and 94.8% (AUROC) of the time, respectively. Testing levels of these three immunological factors may be a valuable tool at the point of care for managing and preventing outbreaks. Further, stem-cell therapy via SCGF-β and M-CSF appear to be promising novel therapeutics for patients with COVID-19.

*(JMIRx Med 2020;1(1):e23582) doi:10.2196/23582*

**KEYWORDS**

infectious disease; SARS-CoV-2; COVID-19; public health; immunity; mass vaccinations; therapeutics; stem-cell growth factor-beta

**Introduction**

Asymptomatic patients who are infected with SARS-CoV-2 have neither clinical symptoms nor abnormal chest imaging. However, these patients have the same infectivity as infected patients with symptoms [1]. Moreover, adult asymptomatic patients have been found to have the same viral loads as symptomatic patients [2]. Studies have shown that age appears to influence whether an infected person is susceptible to illness. Those under the age of 20 years have approximately half the morbidity probability as those over the age of 20 [3]. This improbability of becoming ill from SARS-CoV-2 infection is especially interesting because young children have been found to have 10 to 100 times the viral load as older children and adults, and disproportionately remain asymptomatic [4].

Stem cell growth factor-beta (SCGF-β) has been associated with H7N9 (Asian lineage avian influenza A subtype) and disassociated with H5N1 (a highly pathogenic avian influenza [5,6]). Elevated SCGF-β has also been associated with specific disease states of hepatocellular cancer, Chagas disease, cardiomyopathy, inflammation and insulin resistance, and unstable carotid plaques [7-10]. Interleukin-16 (IL-16), the second most important variable in predicting SARS-CoV-2 immunity or resistance, has been strongly associated with asthma [11].

Prior studies on the biomarkers associated with SARS-CoV-2 immune response and morbidity include interferon-gamma (IFN-γ), interferon-beta (IFN-β), and interleukin-8 (IL-8) [12]. Other previous research on immune parameters associated with SARS-CoV-2 severity and prognosis have involved interleukin-1 beta (IL-1β) and interleukin-6 (IL-6). However, others found reduced immunoglobin G levels in asymptomatic patients [13,14]. The general finding in prior research regarding the pathogen-immune relationship with SARS-CoV-2 is that symptomatic patients have considerably more inflammation and cytokine storm activity than asymptomatic patients [14].

What has been unknown for SARS-CoV-2 are three questions to which the answers are suggested in this study. First, which immunological variables are statistically significant, and how important is each in predicting asymptomatic status? Second, which of those variables, if any, have a strong negative correlation, or relationship, with disease severity (ie, asymptomatic patients’ levels are significantly higher than symptomatic patients)? And third, is there an algorithmic or formulaic model of prognostic biomarkers that can accurately predict morbidity—who will be asymptomatic if infected, and who is at risk of more severe symptoms and disease progression—and why?

**Methods**

This study was based on secondary data published as a supplement in *Nature Medicine* in June 2020 [14]. Therein, immunological factors were measured in 74 patients in the Wanzhou District of China. They were diagnosed as SARS-CoV-2 positive by reverse transcriptase-polymerase chain reaction (RT-PCR) in the 14 days before observations were recorded. The median age of the 37 asymptomatic patients was 41 years (range 8-75 years); 22 were female and 15 were male. For comparison, 37 RT-PCR test–positive patients were selected and matched to the asymptomatic group by age, comorbidities, and sex [14].

In this study, five algorithms, or types, of machine learning—a kind of artificial intelligence employing robust brute-force statistical calculations—were applied to a data set of 74 observations of 34 immunological factors in order to attempt three things: (1) to develop a model to accurately predict which patients will be asymptomatic or symptomatic if infected with SARS-CoV-2; (2) to determine the relative importance of each immunological factor; and (3) to determine if there is any level of a subset of immunological factors that can accurately predict which patients are likely to be immune or resistant to SARS-CoV-2.

Minitab 19, version 19.2020.1 (Minitab LLC), was used to calculate means, 95% CIs, P values, and two-sample t tests of statistical significance. Correlation coefficients were also computed using Minitab via Spearman rho since the data were distributed nonparametrically. A second classification and regression tree (CART) algorithm was also applied in Minitab to cross-validate decision tree results from R in Rattle. Minitab’s CART methodology was initially described by Stanford University and University of California Berkeley researchers in 1984 [15].
The Rattle library, version 5.3.0 (Togaware), in the statistical programming language R, version 3.6.3 (CRAN), was used to apply five machine learning algorithms—a decision tree, extreme gradient boosting (XGBoost), linear logistic model (LLM), random forest, and support vector machine (SVM)—to learn which model, if any, could predict asymptomatic status and how accurately. Rattle randomly partitioned the data to select and train on 80% (n=59), validate on 10% (n=7), and test on 10% (n=7) of observations. Two evaluation methods were used: (1) plots of linear fits of the predicted versus observed categorization; and (2) a pseudo-R² measure calculated as the square root of the correlation between the predicted and observed values. Pseudo-R² measure results were evaluated twice, each using for evaluation data that were held back by being randomly selected during partitioning and averaging the two accuracy findings for the final results.

Rattle’s rpart decision tree was also used to identify if any levels of one or more immunological factors could accurately diagnose someone as asymptomatic (ie, via rules). The decision tree results reported here used 20 and 12 as the minimum number of observations necessary in nodes before the split (ie, minimum split). The trees used 7 and 4 as the minimum number of observations in a leaf node (ie, minimum bucket).

The random forest analysis in Rattle began by running a series of differently sized random forest algorithms, ranging from 50 to 500 decision trees, to learn the optimum number of trees to minimize error. Each random forest consisted of a minimum of six variables, which was closest to the square root of the number of statistically significant variables (ie, 34). The lowest error rate was approximately 200 decision trees.

The five machine learning models and CART classification trees were run, including and excluding SCGF-β to identify if there were alternative prognostic biomarkers and levels in the immune profile that could accurately classify and predict SARS-CoV-2 immunity.

**Results**

In total, 34 of the 53 immunological factors (64.2%) were indicated as statistically significant by P values <.05 from a Spearman rho correlation. Of those 34 factors, 31 were statistically significant with P values <.01. Conversely, 35.9% of the 53 immune factors had no statistically significant association with whether a patient was asymptomatic or symptomatic to SARS-CoV-2.

The 22 factors positively correlated with being symptomatic ranged from a minimum coefficient of 0.205 (monocyte chemotactic protein-3 [MCP-3]) to a maximum of 0.781 (tumor necrosis factor–related apoptosis-inducing ligand [TRAIL]). The 11 factors negatively associated with being symptomatic ranged from a minimum of −.866 (SCGF-β) to a maximum of −0.276 (interferon alpha-2 [IFNα2]) (see Table 1).
### Table 1. Immunological factors associated with SARS-CoV-2 morbidity ranked by Spearman correlation coefficients with 95% CIs and P values.

<table>
<thead>
<tr>
<th>Immunological factor</th>
<th>Abbreviation</th>
<th>Pairwise Spearman correlation to asymptomatic (0) or symptomatic (1) status</th>
<th>95% CI</th>
<th>P value (&lt;.05 target)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNF-related apoptosis-inducing ligand</td>
<td>TRAIL</td>
<td>0.781</td>
<td>0.654 to 0.865</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Growth-related oncogene alpha</td>
<td>GRO-α</td>
<td>0.750</td>
<td>0.611 to 0.845</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Macrophage-colony stimulating factor</td>
<td>M-CSF</td>
<td>0.748</td>
<td>0.608 to 0.843</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-6</td>
<td>IL-6</td>
<td>0.705</td>
<td>0.549 to 0.813</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Granulocyte-colony-stimulating factor</td>
<td>G-CSF</td>
<td>0.697</td>
<td>0.539 to 0.808</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-2</td>
<td>IL-2</td>
<td>0.667</td>
<td>0.499 to 0.787</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Nerve growth factor beta</td>
<td>NGF-β</td>
<td>0.651</td>
<td>0.479 to 0.775</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-10</td>
<td>IL-10</td>
<td>0.614</td>
<td>0.431 to 0.748</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Monocyte chemoattractant protein-1</td>
<td>MCP-1</td>
<td>0.594</td>
<td>0.407 to 0.733</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stem-cell factor</td>
<td>SCF</td>
<td>0.586</td>
<td>0.397 to 0.728</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-15</td>
<td>IL-15</td>
<td>0.527</td>
<td>0.325 to 0.683</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-8</td>
<td>IL-8</td>
<td>0.514</td>
<td>0.311 to 0.673</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interferon-gamma</td>
<td>IFN-γ</td>
<td>0.464</td>
<td>0.252 to 0.633</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-7</td>
<td>IL-7</td>
<td>0.454</td>
<td>0.240 to 0.625</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interferon gamma inducible protein-10</td>
<td>INF-γ-IP-10</td>
<td>0.451</td>
<td>0.237 to 0.623</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-18</td>
<td>IL-18</td>
<td>0.438</td>
<td>0.223 to 0.613</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Platelet-derived growth factor BB</td>
<td>PDGF-BB</td>
<td>0.436</td>
<td>0.220 to 0.611</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interleukin-2 receptor alpha</td>
<td>IL-2Rα</td>
<td>0.388</td>
<td>0.166 to 0.572</td>
<td>.001</td>
</tr>
<tr>
<td>Immunoglobulin G (convalescing)</td>
<td>IgG Conv</td>
<td>0.366</td>
<td>0.143 to 0.544</td>
<td>.013</td>
</tr>
<tr>
<td>Monokine-induced by gamma</td>
<td>MIG</td>
<td>0.364</td>
<td>0.140 to 0.552</td>
<td>.013</td>
</tr>
<tr>
<td>Immunoglobulin G (acute)</td>
<td>IgG Acute</td>
<td>0.330</td>
<td>0.103 to 0.524</td>
<td>.004</td>
</tr>
<tr>
<td>Macrophage migration inhibitory factor</td>
<td>MIF</td>
<td>0.237</td>
<td>0.006 to 0.444</td>
<td>.04</td>
</tr>
<tr>
<td>Monocyte chemotactic protein-3</td>
<td>MCP-3</td>
<td>0.205</td>
<td>-0.270 to 0.416</td>
<td>.08</td>
</tr>
<tr>
<td>Vascular endothelial growth factor</td>
<td>VEGF</td>
<td>0.184</td>
<td>-0.048 to 0.397</td>
<td>.12</td>
</tr>
<tr>
<td>N gene</td>
<td>N</td>
<td>0.180</td>
<td>-0.053 to 0.394</td>
<td>.13</td>
</tr>
<tr>
<td>Interleukin-3</td>
<td>IL-3</td>
<td>0.163</td>
<td>-0.070 to 0.379</td>
<td>.17</td>
</tr>
<tr>
<td>Interleukin-12-p40</td>
<td>IL-12(p40)</td>
<td>0.151</td>
<td>-0.082 to 0.368</td>
<td>.20</td>
</tr>
<tr>
<td>Interleukin-9</td>
<td>IL-9</td>
<td>0.149</td>
<td>-0.084 to 0.366</td>
<td>.21</td>
</tr>
<tr>
<td>Interleukin-1 beta</td>
<td>IL-1β</td>
<td>0.125</td>
<td>-0.107 to 0.345</td>
<td>.29</td>
</tr>
<tr>
<td>Days shed virions</td>
<td>Days shed</td>
<td>0.122</td>
<td>-0.110 to 0.342</td>
<td>.30</td>
</tr>
<tr>
<td>Stromal cell-derived factor-1 alpha</td>
<td>SDF-1α</td>
<td>0.098</td>
<td>-0.124 to 0.320</td>
<td>.41</td>
</tr>
<tr>
<td>Interleukin-12-p70</td>
<td>IL-12(p70)</td>
<td>0.083</td>
<td>-0.149 to 0.306</td>
<td>.48</td>
</tr>
<tr>
<td>Interleukin-17</td>
<td>IL-17</td>
<td>0.067</td>
<td>-0.164 to 0.291</td>
<td>.57</td>
</tr>
<tr>
<td>Interleukin-4</td>
<td>IL-4</td>
<td>0.020</td>
<td>-0.210 to 0.247</td>
<td>.87</td>
</tr>
<tr>
<td>Interleukin-13</td>
<td>IL-13</td>
<td>-0.022</td>
<td>-0.249 to 0.208</td>
<td>.86</td>
</tr>
<tr>
<td>Fibroblast growth factor</td>
<td>FGF</td>
<td>-0.078</td>
<td>-0.302 to 0.153</td>
<td>.51</td>
</tr>
<tr>
<td>Regulated upon activation, normal T-cell expressed and secreted</td>
<td>RANTES</td>
<td>-0.085</td>
<td>-0.308 to 0.146</td>
<td>.47</td>
</tr>
<tr>
<td>Immunological factor</td>
<td>Abbreviation</td>
<td>Pairwise Spearman correlation to asymptomatic (0) or symptomatic (1) status</td>
<td>95% CI</td>
<td>P value (&lt;.05 target)</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Macrophage inflammatory protein-1 beta</td>
<td>MIP-1β</td>
<td>-0.109</td>
<td>-0.330 to 0.123</td>
<td>.35b</td>
</tr>
<tr>
<td>ORF1ab gene</td>
<td>ORF1ab</td>
<td>-0.113</td>
<td>-0.334 to 0.119</td>
<td>.34b</td>
</tr>
<tr>
<td>Macrophage inflammatory protein-1 alpha</td>
<td>MIP-1α</td>
<td>-0.138</td>
<td>-0.356 to 0.095</td>
<td>.24b</td>
</tr>
<tr>
<td>Tumor necrosis factor-alpha</td>
<td>TNF-α</td>
<td>-0.168</td>
<td>-0.383 to 0.065</td>
<td>.15b</td>
</tr>
<tr>
<td>Tumor necrosis factor-beta</td>
<td>TNF-β</td>
<td>-0.197</td>
<td>-0.409 to 0.035</td>
<td>.09b</td>
</tr>
<tr>
<td>Interferon alpha-2</td>
<td>IFNα2</td>
<td>-0.276</td>
<td>-0.478 to -0.046</td>
<td>.02c</td>
</tr>
<tr>
<td>Leukemia inhibitory factor</td>
<td>LIF</td>
<td>-0.312</td>
<td>-0.509 to -0.840</td>
<td>.007c</td>
</tr>
<tr>
<td>Interleukin-5</td>
<td>IL-5</td>
<td>-0.316</td>
<td>-0.512 to -0.089</td>
<td>.006c</td>
</tr>
<tr>
<td>Interleukin-1 alpha</td>
<td>IL-1α</td>
<td>-0.332</td>
<td>-0.526 to -0.106</td>
<td>.004c</td>
</tr>
<tr>
<td>Granulocye-macrophage colony-stimulating factor</td>
<td>GM-CSF</td>
<td>-0.359</td>
<td>-0.548 to -0.134</td>
<td>.002c</td>
</tr>
<tr>
<td>Interleukin-1 receptor alpha</td>
<td>IL-1Rα</td>
<td>-0.359</td>
<td>-0.548 to -0.135</td>
<td>.002c</td>
</tr>
<tr>
<td>Eotaxin</td>
<td>Eotaxin</td>
<td>-0.390</td>
<td>-0.576 to -0.169</td>
<td>.001c</td>
</tr>
<tr>
<td>Cutaneous T-cell-attracting chemokine</td>
<td>CTACK</td>
<td>-0.456</td>
<td>-0.627 to -0.243</td>
<td>&lt;.001c</td>
</tr>
<tr>
<td>Hepatocyte growth factor</td>
<td>HGF</td>
<td>-0.594</td>
<td>-0.733 to -0.407</td>
<td>&lt;.001c</td>
</tr>
<tr>
<td>Interleukin-16</td>
<td>IL-16</td>
<td>-0.827</td>
<td>-0.895 to -0.721</td>
<td>&lt;.001c</td>
</tr>
<tr>
<td>Stem-cell growth factor-beta</td>
<td>SCGF-β</td>
<td>-0.866</td>
<td>-0.920 to -0.780</td>
<td>&lt;.001c</td>
</tr>
</tbody>
</table>

aTNF: tumor necrosis factor.
bStatistically insignificant.
cStatistically significant negative correlations.

When SCGF-β was included in the machine learning analysis, two algorithms predicted and classified SARS-CoV-2 immunity or resistance by being asymptomatic with 100% accuracy: a decision tree and XGBoost. When SCGF-β was excluded, a random-forest algorithm predicted and classified SARS-CoV-2 asymptomatic and symptomatic cases with 94.8% AUROC (area under the receiver operating characteristic) curve accuracy (95% CI 90.17%-100%) (see Table 2).

Notably, both the rpart decision trees and CART classification trees independently identified three prognostic biomarkers at specific levels that could classify asymptomatic and symptomatic cases with 95%-100% accuracy. When SCGF-β was included, all asymptomatic cases had levels >127,656.8, while all symptomatic cases had levels <127,656.8 (Figure 1). When SCGF-β was excluded, as a type of contingency analysis to understand prognostic biomarker levels in other factors better, IL-16 accurately classified asymptomatic cases >44.59 and symptomatic cases <44.59 in 90.4% of the cases. In the remaining 9.6% of cases where IL-16 >44.59, all had macrophage colony-stimulating factor (M-CSF) >57.13 (Figure 2).
Table 2. Comparative accuracy of six machine learning algorithms in predicting SARS-CoV-2 asymptomatic status from immunological factors.

<table>
<thead>
<tr>
<th>Machine learning model</th>
<th>Pseudo-$R^2$ (10% evaluation holdback sample 1) (%)</th>
<th>Pseudo-$R^2$ (10% evaluation holdback sample 2) (%)</th>
<th>Average Pseudo-$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>With SCGF-β</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision tree</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>XGBoost</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>GLM (logistic)</td>
<td>100.00</td>
<td>98.89</td>
<td>99.45</td>
</tr>
<tr>
<td>Random forest</td>
<td>99.46</td>
<td>94.83</td>
<td>97.15</td>
</tr>
<tr>
<td>SVM</td>
<td>78.81</td>
<td>96.99</td>
<td>87.90</td>
</tr>
<tr>
<td><strong>Without SCGF-β</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random forest</td>
<td>97.68</td>
<td>91.91</td>
<td>94.80</td>
</tr>
<tr>
<td>GLM (logistic)</td>
<td>100.00</td>
<td>85.96</td>
<td>92.98</td>
</tr>
<tr>
<td>SVM</td>
<td>77.76</td>
<td>89.69</td>
<td>83.73</td>
</tr>
<tr>
<td>XGBoost</td>
<td>99.42</td>
<td>54.27</td>
<td>76.85</td>
</tr>
<tr>
<td>Decision tree</td>
<td>100.00</td>
<td>2.22</td>
<td>51.11</td>
</tr>
</tbody>
</table>

aSCGF-β: stem cell growth factor-beta.
bXGBoost: extreme gradient boosting.
cGLM: generalized linear model.
dSVM: support vector machine.

Figure 1. Classification and regression tree (CART) of the role of stem cell growth factor-beta (SCGF-β) in predicting SARS-CoV-2 morbidity.
Two-sample $t$ tests for the four factors with the highest positive and negative correlation coefficients, interquartile ranges, outliers, and levels between asymptomatic and symptomatic patients that were statistically significant were computed to ordinally rank factors by their correlation coefficients (Figure 3).

A random forest analysis of the most important variables to accurately classify and predict SARS-CoV-2 patients by binary morbidity ordinally ranked the 34 statistically significant factors. Unsurprisingly, SCGF-$\beta$, and IL-16, followed by growth-related oncogene alpha (GRO-$\alpha$) and TRAIL, respectively, were the most critical factors in predicting morbidity (Figure 4).
Figure 3. Two-sample t tests of statistical significance of the difference in means of four leading prognostic biomarkers for asymptomatic or symptomatic SARS-CoV-2. SCGF-β: stem cell growth factor-beta; IL-16: interleukin-16; GRO-α: growth-related oncogene alpha; TRAIL: tumor necrosis factor–related apoptosis-inducing ligand.
Finally, the results suggest that IL-1β, 3, 4, 9, 12, 13, 17, and RANTES (regulated upon activation, normal T-cell expressed and secreted) are of low importance, or comparative irrelevance, in the pathogen-immune relationship and, that SCGF-β, IL-16, HGF, INFα2, LIF, CTACK, IL-1α, Eotaxin, GM-CSF, IL-1Rα, and IL-5 are valuable in models to predict and classify asymptomatic or symptomatic SARS-CoV-2 cases accurately.

**Discussion**

**Principal Findings**

While it has been speculated that stem cells may play a role in SARS-CoV-2 and other zoonoses’ resistance, prior research has focused on different stem cell involvement than SCGF-β [16-18]. Previous research has also established that stem cells can inhibit viral growth by expressing IFN-γ-stimulated genes and have been particularly effective against influenza A H5N1 virus and resulting lung injuries [19,20]. Stem cell therapy has been hypothesized as a treatment for SARS-CoV-2; however, there is no record in the literature specific as to which factors may influence SARS-CoV-2 infections, favorably or unfavorably, or to what degree until now [21].

Researchers have recently found that symptomatic patients generally have a more robust immune response to SARS-CoV-2 infection, culminating in cytokine storms in the worst cases. Conversely, asymptomatic patients have been found to have a weaker immune response [14]. Because infections are causal to immune response, of particular interest in this study were the most impactful immune-related variables that negatively correlated with asymptomatic status (ie, variables that were greater for asymptomatic patients than symptomatic patients) (marked with a superscripted “c” in Table 1).

This paper’s overarching importance is the identification of immunological factors for diagnoses, treatments, and preclinical prophylactic immune-based approaches to SARS-CoV-2 in the first 7 months of a pandemic that experts now opine will last decades [22]. Immunostimulant approaches are especially valuable because, unlike antivirals and vaccines, they may be given later in the course of the disease to optimize outcomes [21].

The primary importance of this work is machine learning algorithmic models that can predict with high accuracy whether someone, once infected, will be asymptomatic or symptomatic from SARS-CoV-2. This knowledge gives clinicians new tools to identify populations in advance who appear to be at higher risk of danger from the virus. Such devices, especially once reproduced in a more extensive study, may also inform policy decisions as to who needs to shelter in place. Finally, because of the scale of this pandemic and practical constraints as to how many vaccination doses can be manufactured and how quickly this can be done, such tools may become valuable in prioritizing vaccine administration to those in greatest need because they have a higher biological and immunological risk.

This work’s secondary importance is a description of the cytokine and chemokine profile that is associated with asymptomatic or symptomatic SARS-CoV-2 infections. It enables a better understanding of the pathogen-immune relationship. These profiles provide insights into the biological pathways critical for SARS-CoV-2 progression.

---

**Figure 4.** Relative importance of immunological variables from random forest analysis in predicting SARS-CoV-2 morbidity.
As one example, stem cell factors secrete multiple factors that regulate immune cells and modulate them to restore tissue homeostasis. These results suggest that higher levels of SCGF-β (stem-cell factor-beta) may better control immune responses to prevent the more robust reactions universally associated so far with highly symptomatic patients and, further, prevent high morbidity and mortality cytokine storms. A better understanding of the pathogen-immune relationship may enable researchers to prevent and treat patients with SARS-CoV-2 infection more effectively with therapeutics currently untested and unused. This knowledge may also extend to similar zoonotic coronaviruses in the future.

The tertiary importance of this work is identifying three immune factors and precise levels that appear to be prognostic biomarkers as to whether someone, once infected with SARS-CoV-2, will be immune or resistant, as demonstrated by being asymptomatic or not. These insights also suggest new candidates for therapeutic research focused on the relatively newly identified and ill-understood SCGF-β and its role in the immunological process.

The quaternary importance of this work is further proof that machine learning methods can accurately and quickly identify critical elements of disease dynamics that accelerate understanding and improve outcomes during pandemics. Moreover, it is an example of how a “dry” data science laboratory can link to clinical or “wet” laboratory science for real-world applications.

**Limitations**

This study has several limitations. First, it is unknown from the data set how many days passed between exposure to the virus and immunological testing, or whether it was universally the same number of days. Second, because immune profiles are temporally sensitive, ideally, several tests would have been taken over several days, which did not occur (R Jankord, PhD, July 22, 2020). Third, immunological signaling and processing are multifactorial and complex. Therefore, it is unclear why SCGF-β levels are categorically high in asymptomatic patients and low in symptomatic patients, or whether they are causal to SARS-CoV-2 response. Fourth, combinatorial and sequential analysis of these immunological elements may be an important future research area to optimize therapeutic research outcomes. Fifth, at least one study in a leading journal, *The Lancet*, found that Chinese SARS-CoV-2 case data may have been misreported by as much as 400% [23]. That study, and much higher case and fatality numbers in over 200 countries, have created distrust and skepticism of SARS-CoV-2–related data originating from China.

Future research could ameliorate these limitations and focus on a more extensive study group to attempt to reproduce the results. Moreover, a prospective case-control study of patients with decreased SCGF-β levels and supplementation that was protective against SARS-CoV-2 severity and symptoms would be invaluable validation.

**Conclusion**

One implication of these findings is that if we can predict the 80% of society who may be immune or resistant to SARS-CoV-2, or asymptomatic, it may profoundly impact public health intervention decisions as to who needs to be protected and by how much. If, for example, 80% of the shelter-in-place orders and the resultant dramatic reduction in economic and social activity could have been prevented by accurately predicting who is at low risk of infection, the economic benefits alone may have been valued in US$ trillions. The second implication of these findings is evidence that elevated levels of SCGF-β, IL-16, and M-CSF may have a causal relationship with SARS-CoV-2 immunity or resistance, and may have utility as diagnostic determinants to (1) inform public health policy decisions to prioritize and reduce shelter-in-place orders to minimize economic and social impacts; (2) advance therapeutic research; and (3) prioritize vaccine distribution to benefit those with the greatest need and risks first.

**Acknowledgments**

The author wishes to thank Dr Luka Fajs for his helpful comments on drafts of this paper.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

- AUROC: area under the receiver operating characteristic
- GRO-α: growth-related oncogene alpha
- IFN-β: interferon-beta
- IFN-γ: interferon-gamma
- IFN-γR2: interferon-alpha-2
IL-1β: interleukin-1 beta
IL-6: interleukin-6
IL-8: interleukin-8
IL-16: interleukin-16
LLM: linear logistic model
M-CSF: macrophage colony-stimulating factor
MCP-3: monocyte chemotactic protein-3
RANTES: regulated upon activation, normal T-cell expressed and secreted
RT-PCR: reverse transcriptase-polymerase chain reaction
SCF-β: stem-cell factor-beta
SCGF-β: stem cell growth factor-beta
SVM: support vector machine
TRAIL: tumor necrosis factor–related apoptosis-inducing ligand
XGBoost: extreme gradient boosting
Predicting Health Disparities in Regions at Risk of Severe Illness to Inform Health Care Resource Allocation During Pandemics: Observational Study

Tara Fusillo
John F Kennedy High School, Bellmore, NY, United States

Corresponding Author:
Tara Fusillo
John F Kennedy High School
3000 Bellmore Ave
Bellmore, NY, 11710
United States
Phone: 1 516 992 1400
Fax: 1 516 992 1400
Email: ifusil@optonline.net

Related Articles:
Companion article: https://preprints.jmir.org/preprint/22470
Companion article: https://med.jmirx.org/2020/1/e25572
Companion article: https://med.jmirx.org/2020/1/e25573

Abstract

Background: Pandemics including COVID-19 have disproportionately affected socioeconomically vulnerable populations.

Objective: Our objective was to create a repeatable modeling process to identify regional population centers with pandemic vulnerability.

Methods: Using readily available COVID-19 and socioeconomic variable data sets, we used stepwise linear regression techniques to build predictive models during the early days of the COVID-19 pandemic. The models were validated later in the pandemic timeline using actual COVID-19 mortality rates in high population density states. The mean sample size was 43 and ranged from 8 (Connecticut) to 82 (Michigan).

Results: The New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania models provided the strongest predictions of top counties in densely populated states with a high likelihood of disproportionate COVID-19 mortality rates. For all of these models, P values were less than .05.

Conclusions: The models have been shared with the Department of Health Commissioners of each of these states with strong model predictions as input into a much needed “pandemic playbook” for local health care agencies in allocating medical testing and treatment resources. We have also confirmed the utility of our models with pharmaceutical companies for use in decisions pertaining to vaccine trial and distribution locations.

(JMIRx Med 2020;1(1):e22470) doi:10.2196/22470

KEYWORDS

coronavirus; SARS-CoV-2; COVID-19; pandemic; socioeconomic status; predictive model; health care resource allocation
Introduction

Socioeconomic vulnerability can directly influence the severity of pandemics and their impact on mortalities in ways like access to health care, household overcrowding, and comorbidities. Prior studies of swine flu (H1N1) have pointed to these factors as contributors to the spread and severity of that pandemic [1]. Other studies have identified national level correlations that are helpful, but not actionable at a local level where actual health care resource allocation decisions are made [2].

Early and accurate decisioning for health care resource allocations are particularly critical in geographic locations with high population density. This research sought to create a repeatable modeling process that uses readily available data sources to identify the top counties in densely populated states with a high likelihood of disproportionate COVID-19 mortality rates.

Stepwise linear regression was used as the modeling technique. Other similar epidemiological research has also used the stepwise linear regression approach including Thomson et al’s [3] 2006 research on environmental models to predict meningitis epidemics in Africa; Chung et al’s [4] 2012 study of the West Nile encephalitis epidemic in Dallas, Texas; Fulton et al’s [5] 2019 predictive models for hospital-based back surgery demand, and Yu et al’s [6] 2005 study on SARS (severe acute respiratory syndrome).

Our objective was to create a repeatable modeling process to identify regional population centers with pandemic vulnerability.

Methods

Exploratory data research at a national level was performed using county level data (Federal Information Processing Standards [FIPS] for county identification). COVID-19 mortality data sets (deaths per 100,000 people) were created using the Johns Hopkins Dataset [7] and data from the US Census Bureau [8]. Socioeconomic vulnerability data sets at the county level were created using subcomponents of the Centers for Disease Control and Prevention’s (CDC) social vulnerability index (SVI) [9]. The full list of subcomponents can be found in Multimedia Appendix 1.

Scatterplots and trendlines were used to identify variables most correlated with COVID-19 mortalities (see samples in Figure 1). Few, if any, social vulnerability variables correlated across all of the 3142 FIPS counties, but minority status correlated strongly in certain regions, particularly those with high mortality rates. These initial findings led the author to focus the next phase of research and modeling on state level rather than national level correlations. County-level dependent variable data sets were created using COVID-19 mortality and population data from the Corona Data Scraper website (data service that scrapes county level COVID-19 data on a daily basis) [10] as well as from USAFacts [11] for cumulative mortalities as of April 8, 2020, and May 8, 2020, respectively.

Figure 1. Examples of regional scatterplots and trendlines. CDC: Centers for Disease Control and Prevention.

County-level independent variable data sets were created using socioeconomic data from the County Health Rankings website, a collaboration data set created by the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute [12]. Given the year to year stability of most of these socioeconomic variables, the latest data available from the County Health Rankings website was used, and no attempt was made to augment the data set to try to match the time-series to either April 8, 2020, or May 8, 2020. The full list of independent variables available in this data set can be found in Multimedia Appendix 1.

Cumulative COVID-19–specific mortality data (deaths per 100,000 people) by county for states with a high mortality rate (New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania) as of May 8 was used as the dependent variable [13]. The full ranking of states by mortality rate can be found in Multimedia Appendix 1.

The May 8 date was used to ensure that the dependent variable would be tuned to a timeframe at or around the peak in daily mortalities when health care resources (testing, treatment, and tracing) were typically most needed. The mortality curves in Figure 2 provide support for May 8 as the overall date for...
mortality predictions as shown in the Institute for Health Metrics and Evaluation data set [14].

A stepwise linear regression technique was used to build each state level model. All relevant independent variables were initially used in the model (ie, include severe housing problems, but exclude violent deaths). Next, the variable with the lowest T-statistic was removed and the linear regression was rerun. This process was repeated until all T-statistics for the remaining independent variables were near a value of 2 or greater.

**Figure 2.** Examples of COVID-19 mortality curves from the Institute for Health Metrics and Evaluation [14].
Results

Predictive models were completed for New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania, with statistically significant results. The final list of model variables, coefficients, variable correlations, sample sizes, and $P$ values can be found in Multimedia Appendix 1.

Model validation comparing predicted to actual county rankings by cumulative mortality rate (deaths per 100,000 people) through May 8 (see Tables 1 and 2 for New York and New Jersey), and data visualizations using state-level maps were completed (Figure 3). These validations were used to share model methods and results with each state’s Department of Health Commissioner and, where appropriate, with an outside agency.

Table 1. Top counties by mortality rate for New York (actuals vs model prediction). Both actual values and modeled predictions indicate that these counties had high mortality rates relative to other counties within the state.

<table>
<thead>
<tr>
<th>New York county</th>
<th>Deaths per 100,000 people (5/8 actuals), n</th>
<th>Deaths per 100,000 people (5/8 model), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bronx</td>
<td>198</td>
<td>197</td>
</tr>
<tr>
<td>Kings</td>
<td>167</td>
<td>155</td>
</tr>
<tr>
<td>Queens</td>
<td>188</td>
<td>148</td>
</tr>
<tr>
<td>Westchester</td>
<td>115</td>
<td>118</td>
</tr>
<tr>
<td>New York</td>
<td>104</td>
<td>118</td>
</tr>
<tr>
<td>Rockland</td>
<td>129</td>
<td>114</td>
</tr>
<tr>
<td>Nassau</td>
<td>134</td>
<td>102</td>
</tr>
<tr>
<td>Suffolk</td>
<td>87</td>
<td>91</td>
</tr>
<tr>
<td>Richmond</td>
<td>128</td>
<td>88</td>
</tr>
<tr>
<td>Orange</td>
<td>68</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 2. Top counties by mortality rate for New Jersey (actuals vs model prediction). Both actual values and modeled predictions indicate that most of these counties had high mortality rates relative to other counties within the state (exception marked with "a").

<table>
<thead>
<tr>
<th>New Jersey county</th>
<th>Deaths per 100,000 people (5/8 actuals), n</th>
<th>Deaths per 100,000 people (5/8 model), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hudson</td>
<td>140</td>
<td>139</td>
</tr>
<tr>
<td>Passaic</td>
<td>143</td>
<td>126</td>
</tr>
<tr>
<td>Bergen</td>
<td>143</td>
<td>121</td>
</tr>
<tr>
<td>Union</td>
<td>152</td>
<td>116</td>
</tr>
<tr>
<td>Essex</td>
<td>175</td>
<td>101</td>
</tr>
<tr>
<td>Middlesex</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>Hunterdon</td>
<td>35a</td>
<td>93</td>
</tr>
<tr>
<td>Somerset</td>
<td>100</td>
<td>88</td>
</tr>
<tr>
<td>Morris</td>
<td>103</td>
<td>88</td>
</tr>
</tbody>
</table>

*a*Indicates low mortality rate relative to all other counties within the state.
Four further validations were completed. The first validation was to check model performance using COVID-19 mortality data on April 8, 2020, instead of May 8, 2020. This validation tested whether models using data available early in the pandemic would have been sufficient to make accurate predictions. The same variables were used, but coefficients were recalibrated with the April 8 data set. The April 8 and May 8 model outputs were compared to test for stability in the top counties predicted for high COVID-19 mortality rates. For New York, New Jersey, and Connecticut, the models proved to be stable. For Massachusetts, the 4/8 model performance was not stable, but this was easily corrected by using case data in the place of mortality data. This is an important finding as it validates the predictive power contained in early case data, which is more readily available at the start of a pandemic. The New York and Massachusetts model validation result summaries can be found in Multimedia Appendix 1.

The second validation was to check model performance beyond May 8, 2020 (ie, using the May 8 model to predict July 31 mortalities). Results were less stable as most states began their reopenings in mid-May creating differential effects by county. However, the models for New York, New Jersey, Connecticut, and Massachusetts continued to identify the counties with the highest cumulative mortality rates.

The third validation was to check the independent variables for multicollinearity, with a specific focus on the correlations between “Black” and variables such as “severe housing” and “uninsured.” Strong multicollinearity was seen in Connecticut, Massachusetts, Louisiana, and Michigan partially explaining why these variables did not remain in the model after the stepwise regression process. Multicollinearity results for these states are presented in Multimedia Appendix 1. Future models could consider composite variables to address this multicollinearity and to maintain combined effects such as “Black,” “severe housing,” and “uninsured.”

The fourth validation leveraged an out-of-sample methodology. For New York, New Jersey, Connecticut, and Massachusetts, only one half of the data points (ie, half of the counties in each state) were used to build the model. Coefficients were recalibrated and variables were removed if T values fell below 2. In each state, the out-of-sample model continued to identify the top counties for cumulative mortality rates through May 8.

With models and validations completed, the Departments of Health for New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania were contacted (Table 3). Additionally, agents of the New York Department of Health (Northwell Health and CORE), a third-party statistical modeling firm for Connecticut (COVIDACTNOW), and a third-party modeling firm for Pennsylvania (Mathematica) were contacted. Response from these contacts were positive and, in some cases, occurred within an hour of outreach (Northwell Health). This response indicates the strong need for this type of health care resource allocation tool for pandemics and other health crises. In fact, the Pennsylvania Department of Health indicated this tool’s importance in a second wave of COVID-19.

The final data sets for our New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania models are contained in Multimedia Appendix 1.
Table 3. State Department of Health (DOH) contact summary.

<table>
<thead>
<tr>
<th>DOH or outside agency</th>
<th>Models shared</th>
<th>Receipt accepted</th>
<th>Zoom session</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York DOH</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Northwell Health (New York)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CORE (New York)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey DOH</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut DOH</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVIDACTNOW (Connecticut)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts DOH</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michigan DOH</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louisiana DOH</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania DOH</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mathematica</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

The research described in this paper has shown the extent to which Black Americans, people living in crowded housing units, and households with less access to health care are at higher risk of severe illness during a pandemic. The results of other studies provide several possible explanations for these findings. Individuals with lower income are more likely to live in crowded housing units and multifamily homes [15]. Lower income households are also structurally disadvantaged in their access to medical insurance and health care [16]. Some studies have also pointed to a concept called “weathering” within the Black American community. Arline Geronimus, a professor of public health at the University of Michigan, showed through her research that among Black communities, coping with financial strain, discrimination, and barriers to good education elevates the stress response, contributing to obesity, diabetes, hypertension, and heart disease [17].

The CDC conducted research on the co-occurrence of COVID-19 and certain ethnicities using a sample of 580 patients with lab-confirmed data [18]. Their results showed more hospitalizations in Black patients than White patients. The researchers cited underlying medical conditions, work circumstances, and living conditions to be major factors in COVID-19 mortalities in their sample. For example, members of racial and ethnic minorities were more likely to live in densely populated areas, making it more difficult to practice social distancing and being more susceptible to contracting and spreading COVID-19. These members also lived farther away from grocery stores and medical facilities, thus being less able to receive necessary resources and medical attention. Other examples cited included Hispanic and Black American workers employed in higher-risk industries and often lacking paid sick leave. The researchers also hypothesized that these types of workers were more likely to continue working despite being sick, thus exposing other workers to the disease. The CDC recommended at the conclusion of this research that public health officials communicate to different population groups about COVID-19 and provide more health care services to ethnic minority groups.

While such studies are insightful, we are not aware of any research that translates these impacts into predictive models that can be used to direct local health care resources to the communities most likely to need them, thereby reducing mortalities caused by an ongoing set of institutional inequities.

That said, some organizations have attempted to create health care resource allocation methods using descriptive statistics. The CDC created the SVI, allowing health care communities to see which factors contribute to socioeconomic vulnerability. The CDC SVI factors are grouped into four groups: Socioeconomic Status, Household Composition & Disability, Minority Status & Language, and Housing Type & Transportation. Although these factors are crucial inputs in identifying specific vulnerable communities, these four categories of factors alone are not sufficient to create the types of predictive models that state and local health care agencies can use. One example of this insufficiency is a recent study at Emory University where researchers identified correlations between COVID-19 mortalities and the SVI at a national level in the United States [2]. While this study is valuable, it stopped short of recommending methods or processes to effectively distribute health care resources to specific counties in the United States, particularly in the early days of the pandemic. Another study from the Surgo Foundation stated that COVID-19 created new challenges for many communities tied to health and structural factors that were not completely captured by the CDC SVI [19]. In addition to the four socioeconomic factors provided by the CDC, the Surgo Foundation added two more factors: Epidemiologic Factors and Healthcare System Factors. They stated that underlying health conditions in addition to health care system factors have been proven to greatly increase a community’s vulnerability during a pandemic. The Surgo Foundation combined these two factors with the CDC SVI to create the COVID-19 Community Vulnerability Index (CCVI). The Surgo Foundation created heatmaps to show retrospectively which counties were most vulnerable as measured by their CCVI. Similar to the Emory University study, however, this methodology did not create a predictive model to identify where
the mortalities would be highest at peak periods in a pandemic. We also compared the Surgo Foundation heatmap to our own predictive model rankings and confirmed that our projections of the top counties by per capita mortalities were far closer to actual peaks. Results of this comparison are shown in Multimedia Appendix 1.

During the early phases of the research described in this paper, we used the CDC’s SVI, similar to the Emory University study. We explored all of the subcategory factors in the SVI, but none showed strong correlations at a FIPS county level across the United States. We then grouped states together by region and found strong correlations in the most densely populated regions, particularly with the minority status subfactors. Similar to the Surgo Foundation study, we posited that the CDC’s SVI subfactors alone would be insufficient to build predictive models, so a far more complete independent variable data set of socioeconomic and health care data was sourced from the County Health Rankings website as discussed earlier. The key differentiation of our work is the predictive models for each state given the local differences in how each of the factors act as predictors of peak COVID-19 per capita mortalities. Our experience in meeting with the Pennsylvania Department of Health in early June 2020 confirms the uniqueness and usefulness of the approach given that they will be using the predictive modeling process for health care resource allocations in the event of a potential second wave of COVID-19 in Fall 2020. Other state-level departments of health and agents of these governmental functions were similarly intrigued by our predictive approach including those in New York, New Jersey, and Connecticut. Finally, the Chief Information Officer of Johnson & Johnson has forwarded the author’s research to J&J’s Health and Human Services Group for possible use in decisions pertaining to vaccine trial and distribution locations.

Limitations

A number of limitations must be acknowledged. The models employed in the analyses are reliant on the accuracy of the data sets compiled. COVID-19 mortality data in particular has been notoriously difficult for states to report accurately at a county level throughout the pandemic for reasons including mortality cause classification errors at the offices of the local coroner [20]. This systemic undercounting could have created some correlations between our variables and reporting errors. That said, if reporting errors are similar across counties within a state, then these unwanted effects to our model are likely to be small since we created a different model for each state. The models are only valid within the range of county-level data for the following states: New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania. Models would need to be rebuilt and validated for each additional state. In addition, models for states with lower levels of per capita mortalities were far less predictive, although health care resource allocations would be less critical in those regions.

In addition, we would note that while we did not find significant correlations between the subcomponents of the CDC SVI and COVID-19 mortalities, other data sets at the FIPS level (eg, American Community Survey data and census data) might yield different results. For example, there may be variables that are correlated with one another that also correlate with COVID-19 mortalities. Gore et al [21] showed the difficulties in teasing apart specific population demographic measures at a granular level into a linear regression model since so many of these variables correlate highly with one another.

Conclusions

Our modeling process can be used for the early identification of the communities most in need of health care resources during future pandemics or health crises. The COVID-19 models and the overall methodology have been received with enthusiasm by the New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan, and Pennsylvania Departments of Health. All Commissioners responded positively, which validates our hypothesis that neither other researchers nor the Departments of Health themselves have developed a similar modeling process as a means to allocate scarce health care resources such as testing, treatment, tracing, education, and communication. We have also confirmed the utility of our models with pharmaceutical companies for use in vaccine trial and distribution location decisions.

Future modeling processes could also include building hierarchical models to improve county rankings by better accounting for effects of clustered variables similar to Fulton et al [5] in their 2019 models for predicting hospital-based back surgery by geography. The gradient boosting approach leveraged by Fulton et al [5] may also be useful to examine states with lower population densities where our stepwise linear regression models proved to be weaker. Finally, the group-personalized regression approach pioneered by Palmius et al [22] in their 2018 models for predicting mental health scores by group rather than for an overall population could also be explored.

Other research papers evaluated during the course of this research and other data sets referenced in this research can be found in Multimedia Appendix 2.
Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.

[PDF File (Adobe PDF File), 1367 KB - med_v1i1e22470_app1.pdf]

Multimedia Appendix 2
Other data sets referenced in this research and research papers evaluated during the course of this research.

[DOCX File, 18 KB - med_v1i1e22470_app2.docx]

References


10. Corona Data Scraper. Davis L. URL: https://coronadatascraper.com/ [accessed 2020-08-10]


12. How Healthy is your County? County Health Rankings. URL: https://www.countyhealthrankings.org/ [accessed 2020-08-10]


Abbreviations

CCVI: COVID-19 Community Vulnerability Index
CDC: Centers for Disease Control and Prevention
FIPS: Federal Information Processing Standards
SARS: severe acute respiratory syndrome
SVI: social vulnerability index

©Tara Fusillo. Originally published in JMIRx Med (https://med.jmirx.org), 02.12.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://med.jmirx.org/, as well as this copyright and license information must be included.