JMIRx Med

Overlay journal for preprints with post-review manuscript marketplace
Volume 2 (2021), Issue 1 ISSN 2563-6316 Editor in Chief: Edward Meinert, MA (Oxon), MSc, MBA, MPA, PhD, CEng, FBCS, EUR ING

Contents

Peer-Review Reports

Peer Review Reports

Authors' Response to Peer Reviewss

Short Paper

Original Papers

Peer-Review Report

Peer Review of “Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis”

Anonymous

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22195

Companion article: https://med.jmirx.org/2021/1/e27258/

Companion article: https://med.jmirx.org/2021/1/e22195/

(JMIRx Med 2021;2(1):e27103) doi:10.2196/27103

KEYWORDS

infectious diseases; testing; population density; policy; coronavirus; COVID-19; SARS-CoV-2

This is a peer review submitted for the paper “Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis.”

Round 1 Review

General Comments

This paper [1] signals the need for a more nuanced COVID-19 testing strategy. The authors propose using population density–driven testing to help address this need. Testing strategies certainly have room for improvement and continuous assessment, especially in emergent situations like COVID-19. Maps are great visualization tools.

Specific Comments

Major Comments

This paper communicates that adjusting testing strategies by population density will save lives and livelihoods. While I think there is merit to finding effective ways to account for population density, especially in contexts with high-quality census and robust public health surveillance data, there is a host of other dynamic factors that play into the complicated pathway between population density, testing, and saving lives and livelihoods that are not accounted for in the current version of this paper.

This draft also uses absolute terms and expressions that do not seem appropriate given the scope of the study. The authors might benefit from speaking in less absolute terms, remove anecdotal examples such as the elevator vs football field in exchange for more standardized epidemiological measures, and include in the paper a discussion about the limitations of using their proposed population density–driven testing. The paper should also speak more to the nature (eg, challenges) of public health data, monitoring and surveillance, and the role of testing in this context. As a policy-oriented paper, it should also discuss more of the potential impacts of modifying a testing strategy (pros and cons), including the costs associated with changing the current testing strategy. The paper might also want to address whether or not adjusted testing strategies based on population density (or similar measures) have successfully been done elsewhere.

Conflicts of Interest

None declared.

Reference

1. Budhwani KI, Budhwani H, Podbielski B. Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis. JMIRx Med 2021;2(1):e22195. [doi: 10.2196/22195]

Edited by G Eysenbach; submitted 11.01.21; this is a non–peer-reviewed article;accepted 11.01.21; published 03.02.21.

Please cite as:
Anonymous
Peer Review of "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis"
JMIRx Med 2021;2(1):e27103
URL: https://med.jmirx.org/2021/1/e27103/
doi:10.2196/27103
PMID:

Peer-Review Report

Peer Review of “Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis”

Ayman A Allam[1], MD

Faculty of Medicine, Zagazig University, Zagazig, Egypt

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22195

Companion article: https://med.jmirx.org/2021/1/e27258/

Companion article: https://med.jmirx.org/2021/1/e22195/

(JMIRx Med 2021;2(1):e27257) doi:10.2196/27257

KEYWORDS

infectious diseases; COVID-19; SARS-CoV2; coronarvirus

This is a peer review submitted for the paper “Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis.”

Round 1 Review

General Comments

In this paper [1], the authors prospectively analyzed COVID-19 data obtained from 67 Alabama counties using testing realignment along population density instead of density agnostic per capita. They concluded that adjusting the distribution of testing capacity to also account for population density will improve monitoring and response to blunt the speed and spread of the virus.

Generally, the manuscript is properly structured and well understood.

Specific Comments

Minor Comments

1. Change the subtitle “Policy Proposal” to “Introduction” or “Background.”

Conflicts of Interest

None declared.

Reference

1. Budhwani KI, Budhwani H, Podbielski B. Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis. JMIRx Med 2021;2(1):e22195. [doi: 10.2196/22195]

Edited by E Meinert; submitted 18.01.21; this is a non–peer-reviewed article;accepted 18.01.21; published 03.02.21.

Please cite as:
Allam AA
Peer Review of “Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis”
JMIRx Med 2021;2(1):e27257
URL: https://med.jmirx.org/2021/1/e27257/
doi:10.2196/27257
PMID:

Peer-Review Report

Peer Review of “A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study”

Mo Salman[1], BVMS, MPVM, PhD, DACVPM, FACE

College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, United States

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22617

Companion article: https://med.jmirx.org/2021/1/e27537

Companion article: https://med.jmirx.org/2021/1/e22617

(JMIRx Med 2021;2(1):e27260) doi:10.2196/27260

KEYWORDS

COVID-19; pandemics; infection control; models; experimental; longitudinal studies; statistical modeling; epidemiology

This is a peer review submitted for the paper “A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study.”

Round 1 Review

General Comments

It seems that the aim of this submission [1] is to report a study conducted to show an approach for normalization epidemic curves from various countries using retrospective data, particularly from the city of Rio de Janeiro. The submission lacks a recognized structure to present a study with its aim and details of data sources. Furthermore, the submission includes some terms that are not appropriate for describing infectious disease in a population such as contamination and contamination cycle instead of exposure and infection rates.

Specific Comments

1. The aim of the study should be stated in a precise statement with supportive ways to test the underlying hypothesis;
2. Details of the analytical approach should be given with its assumptions and limitations;
3. Sources of the data with overall reliability can be detailed;
4. Use the appropriate and conventional terms of infectious diseases by checking the contents of the submission with reliable epidemiologists.

Round 2 Review

I am satisfied with the modifications to the new version. Almost all of my concerns were addressed in the new version. I will let the readers decide about the validity of the model since the authors elaborated on the approach.

Conflicts of Interest

None declared.

Reference

1. De Carvalho EA, De Carvalho RA. A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study. JMIRx Med 2021 Mar 18;2(1):e22617 [FREE Full text] [doi: 10.2196/22617]

Edited by E Meinert; submitted 18.01.21; this is a non–peer-reviewed article;accepted 27.01.21; published 18.03.21.

Please cite as:
Salman M
Peer Review of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study"
JMIRx Med 2021;2(1):e27260
URL: https://xmed.jmir.org/2021/1/e27260
doi:10.2196/27260
PMID:

Peer-Review Report

Peer Review of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study"

Shelley N Facente[1], PhD

United States

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22617

Companion article: https://med.jmirx.org/2020/1/e27537/

Companion article: https://med.jmirx.org/2021/1/e22617

(JMIRx Med 2021;2(1):e27536) doi:10.2196/27536

KEYWORDS

COVID-19; pandemics; infection control; models; experimental; longitudinal studies; statistical modeling; epidemiology

This is a peer review submitted for the paper "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study."

Round 1 Review

General Comments

This paper [1] attempts to make some interesting comparisons in epidemic curves but is far too long, requires a thorough copy edit by a native English speaker, and does not appear to be very scientific in nature—it would require a substantial overhaul in both methods and write-up to be suitable for publication in this journal.

Specific Comments

Major Comments

1. I think the paper would benefit from focusing on the central point, with only a minor description of any methods necessary to make that point. For example, most of section 2 ("On the Nature of the Observed Data") includes a lengthy description of the use of moving averages and R^2 values, which readers can be expected to understand (and those who do not could learn via appropriate citations). Certainly, the formula for R^2 is not necessary to include. These types of revisions would make the paper a more readable length.
2. The authors claim the Chinese epidemic cycle has a Gaussian shape, but even in their own drawing it appears much more like a gamma distribution (as would be expected) than Gaussian. They may want to reconsider this section.
3. The crux of this paper seems to be looking for epidemic curves that appear similarly shaped in various countries, and using that to predict the curves for countries where the cycle has not yet reached what the authors term the "MLCE end." I have never heard of epidemic predictions using a method such as this, and it seems there are many far superior methods for estimating trajectory, so this is a curious choice.
4. Citations require appropriate formatting for journal publications.

Conflicts of Interest

None declared.

Reference

1. De Carvalho EA, De Carvalho RA. A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study. JMIRx Med 2021 Mar 18;2(1):e22617 [FREE Full text] [doi: 10.2196/22617]

Edited by E Meinert; submitted 27.01.21; this is a non–peer-reviewed article;accepted 27.01.21; published 18.03.21.

Please cite as:
Facente SN
Peer Review of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study"
JMIRx Med 2021;2(1):e27536
URL: https://xmed.jmir.org/2021/1/e27536
doi:10.2196/27536
PMID:

Peer-Review Report

Peer Review of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study"

Jorge Andrés Delgado-Ron[1], MD, MPPGA

Health and Community Design Lab, School of Population and Public Health, The University of British Columbia, Vancouver, BC, Canada

Related Articles:

Companion article: https://preprints.jmir.org/preprint/25610

Companion article: https://med.jmirx.org/2021/1/e28718/

Companion article: https://med.jmirx.org/2021/1/e25610/

(JMIRx Med 2021;2(1):e28714) doi:10.2196/28714

KEYWORDS

public health; global health; COVID-19; hypertension; risk; strategy; mental health; behavior; response; anxiety; vaccine; retrospective; perception; prevention; intention

This is a peer review submitted for the paper "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study."

Round 1 Review

General Comments

This study [1] aimed to investigate whether Australians with hypertension have higher risk perceptions, anxiety, and prevention intentions than Australians without hypertension during COVID-19 restrictions in April and June 2020. The authors used a national survey subsample (those who reported hypertension and not other comorbidities). They matched them with controls using age, gender, education, and health literacy. This is a nationally representative sample that includes several dimensions of an individual's mental health. The question is relevant for future public health interventions.

Overall, the study has several weaknesses and does not appropriately answer the study aim because the reported results are not consistent with the proposed methods. The authors also failed to address alternative explanations to their findings. Please see my detailed feedback after the minor comments.

Specific Comments

Major Comments

1. There is a major disconnect between the proposed methods and the results. Moreover, the authors need to clarify the assumptions that led to the selection of their methods.

2. The overall organization can improve. Some methods are presented in the *Results* or *Discussion* section, and some discussion points are introduced in the *Results* section.

3. The authors need to rewrite the *Introduction* section to better contextualize the potential mediators between exposure and outcome with relevant literature.

4. The authors need to rewrite the discussion emphasizing their findings and addressing their limitations and alternative explanations to their study results.

Minor Comments

1. The tables need to be reworked to not confuse multiple regression and marginal mean difference (MMD).

2. Tables are stand-alone pieces. Some of the methodologies need to be incorporated as a footnote.

3. Some typos need to be fixed across the manuscript.

4. Ethics need to be clarified (not a main concern as this is a secondary analysis).

Detailed Feedback:

Title/Abstract and References

1. Ideally, the title needs to include the study design, the population (Australia), and the study's specific outcomes. Please consider changing it to better reflect your primary exposure: hypertension (eg, "The Impact of Hypertension on Adults' Anxiety During COVID-19 Restrictions").

2. The paper has relatively few references (15); some are press articles (3). The authors could strengthen their writing by considering some of these references:

- https://doi.org/10.1093/eurpub/cky114
- https://doi.org/10.1586/14760584.2015.964212
- https://apps.who.int/iris/bitstream/handle/10665/251671/WHO-HIS-TTi-GAP-16.2-eng.pdf

Introduction

1. The introduction is just one paragraph long. It discusses why hypertensive people could experience increased levels of COVID-19–related anxiety. However, it misses critical points at the center of this debate during the pandemic's early stages (time of the survey). For instance, the role of antihypertensive medication as a potential risk factor on those infected by SARS-CoV-2:

- https://pubmed.ncbi.nlm.nih.gov/32737124/
- https://doi.org/10.1056/NEJMoa2007621

and existing studies on risk perception among people with chronic disease:

- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0237296

2. The research question would be clear and justified if the points considered above are included. I suggest adding details about the population (country).

Methods

1. The data selection process is clear *after* one reads the whole paper but not after reading the *Methods* section. I suggest mentioning early on that subjects with additional comorbidities were excluded from the sample. There is no mention of the matching method used and whether this was done manually or automatically ("randomly matched" is mentioned, but what type of randomization was used?). I would also add a line about (a) why you selected these covariates and (b) the test used to assess an adequate balance between the matched pairs.

2. There is no mention of ethics approval for this study. I understand the original survey was approved by the University of Sydney Human Research Ethics Committee (2020/212). Please mention whether this study is covered under the same authorization.

3. There is no mention of the absence or presence of systematic differences between the followed-up sample and those who decided not to participate for a second time. Was this tested? If there are differences, what are the potential implications?

4. Exposure: Please mention the definition of exposure in the *Methods* (self-reported).

5. Covariates: The *Methods* section reports using the health literacy single-item screener and the Consumer Health Activation Index patient activation measure. I understand these are validated tools. Please add a line about what these tools measure and why they are relevant to the current analysis.

6. Outcomes: Please detail more about risk perceptions and prevention behaviors in the *Methods* section.

7. Statistical analysis: (a) The use of "linear models for continuous outcomes, generalized linear models with modified Poisson approach for dichotomous outcomes, [and] ordinal logistic regression for ordered categorical outcomes" is mentioned. However, maximum mean discrepancy is reported. This method was not described in the appropriate section.

(b) An explanation as to why a modified Poisson approach was used instead of a logistic or log-binomial regression is needed. Similarly, the *Results* section shows an adjusted relative risk. However, this is a cross-sectional sample. The use of relative risk needs to be justified.

8. Data availability: Consider mentioning something regarding data availability.

Results

1. Tables are supposed to be stand-alone. Please add a footnote to Table 1 indicating your matching methodology. Consider adding the standardized mean difference to check the balance between cases and controls. Please tell the reader what you meant by the social distancing score scale. Please explain what is meant by patient activation. Please indicate whether the prescription is specific to hypertension.

2. Consider adding a supplementary table with the results from the follow-up period.

3. Table 2 results are not consistent with the proposed methods nor with the title of the table. Regression models result in exponentiated coefficients presented as odds ratios. In contrast, Table 2 shows MMD (or "MDD" for the social distancing score). Please present your MMD distributions in a separate table (or in the text) and introduce the appropriate methods in the previous section. Consider reporting IQR instead of 95% CI.

4. Please review the following numbers as they do not add up to 1005: "On average the hypertension sample thought that 7% of people who get COVID-19 would die as a result, and 63% would only experience mild symptoms."

5. "On average the mean STAI was 1.90 units higher (95% CI 0.19-3.61, P=.03, Cohen d=0.13) for those with hypertension (40.75) than matched controls (38.85), with both groups higher than normal range, but below clinical levels." The interpretation should be moved to the *Discussion* section. Please explain what you mean by "below clinical levels" as well as your reference scale.

6. Please clarify whether you adjusted for baseline characteristics in these analyses: "At follow-up, there was no longer a significant difference between the hypertension and control groups for influenza vaccination."

Discussion and Conclusions

1. The discussion does not start by stating the study's main findings (the influence of hypertension in the selected outcomes). Instead, it starts by comparing the overall sample with previous results in the same reference population.

2. The results are not discussed from multiple angles. For instance, the authors write, "Those with hypertension were more likely to take up the influenza vaccine during lockdown compared to healthy controls." Could this be an effect of requiring care more often than healthy individuals? Patient activation is different from patient engagement.

3. The authors do not differentiate between willingness to get a vaccine and those who have already gotten a vaccine. Were there active vaccination campaigns between the two survey waves?

4. The authors mention several limitations of the study without detailing why they are limitations and how they were addressed. For instance, the authors write, “The sample was recruited via an online panel and social media, and has a low proportion of culturally and linguistically diverse participants.” What is the implication of this on the interpretation of your results? Did you do something to address such a shortcoming? Also, what are other implications?

5. Are people online more likely to be exposed to news generating anxiety or promoting vaccination? While this is just an example, most limitations lack this broader consideration.

6. Finally, conclusions are overextended and assume a causal effect: “Anxiety was above normal levels for all groups during the COVID-19 lockdown. This was higher in the hypertension group and appeared to translate to higher influenza vaccination intentions”; this is not consistent with the variable measured (intentions + uptake).

Round 2 Review

Specific Comments

The authors have addressed most if not all of the comments. I think the paper needs some proofreading, but that should not prevent its acceptance.

Conflicts of Interest

None declared.

Reference

1. Bonner C, Cvejic E, Ayre J, Isautier J, Semsarian C, Nickel B, et al. The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study. JMIRx Med 2021 Mar 30;2(1):e25610 [FREE Full text] [doi: 10.2196/25610]

Edited by E Meinert; submitted 11.03.21; this is a non–peer-reviewed article;accepted 11.03.21; published 30.03.21.

Please cite as:
Delgado-Ron JA
Peer Review of “The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study”
JMIRx Med 2021;2(1):e28714
URL: https://xmed.jmir.org/2021/1/e28714
doi:10.2196/28714
PMID:

Peer-Review Report

Peer Review of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study"

Dinesh Neupane[1], PhD

Johns Hopkins University, Baltimore, MD, United States

Related Articles:

Companion article: https://preprints.jmir.org/preprint/25610

Companion article: https://med.jmirx.org/2021/1/e28718/

Companion article: https://med.jmirx.org/2021/1/e25610/

(JMIRx Med 2021;2(1):e28717) doi:10.2196/28717

KEYWORDS

public health; global health; COVID-19; hypertension; risk; strategy; mental health; behavior; response; anxiety; vaccine; retrospective; perception; prevention; intention

This is a peer review submitted for the paper "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study."

Round 1 Review

The paper written by Bonner et al [1], describing the impact of COVID-19 restrictions on people with hypertension, provides important information comparing the current status of risk perceptions, anxiety, and prevention intentions among hypertensive patients compared to healthy controls. The paper is well written, the methods are described well, and the results are presented clearly. I think the manuscript will benefit a lot if the authors consider my comments below.

Title

I suggest changing the title so that it is clear and informative and reflects the study's aim and approach. For example: "Risk Perceptions, Anxiety, and Prevention Intentions Among Hypertensive Patient Due to COVID-19 Restrictions in Australia: A Case-Control Study"

Abstract

Conclusion, second line: Who are vulnerable groups? Have you not reported that there is no difference between cases and controls for the majority of outcome variables? Does this not mean mental health screening should be required for all? I suggest the authors revisit the sentence below in the abstract and conclusion: "…may require targeted psychological screening for vulnerable groups."

Methods

1. Is it possible to add a description of the sample size and response rate? This is currently missing in the *Methods* section.

2. Although the reasons for using a linear model, generalized linear model, and ordinal logistic regression are described, it is unclear from the text which test was applied for which estimate. Elaborating on this in the *Methods* section will help readers to understand the methods more appropriately.

Results

1. Table 1: 42% of participants in the control group indicated that they were taking prescription medicine. What type of medicines were they using? Did you not include healthy controls?

2. Risk perception: What statistical test was applied to calculate the MMD coefficient? I hope you have checked the normality assumptions as the data only have a score range of 0-10. I would be cautious to apply linear regression for such types of data.

3. I suggest adding a table (similar to Table 2) for the follow-up results.

Discussion

I think the *Discussion* section can be expanded a little bit. The *Results* section has some salient points that warrant discussion. A few suggestions:

a. Why is only the willingness to get the influenza vaccine significant? What could be the possible reasons?
b. Why is there no statistically significant difference for risk perception? Any literature to support this?
c. I think people with hypertension must be more cautious for adopting preventive measures such as social distancing because their mortality and morbidity are often high. However, the results indicate that people with hypertension also have a similar social distancing score. Is this because the score is too high for both groups, or could there be other potential reasons (eg, the same level of access to preventive

measures/knowledge, lack of awareness that people with underlying conditions have a high level of mortality, etc)?

Limitations

Hypertension often presents with other chronic conditions. Including people with other chronic conditions might produce different results since people with both hypertension and other chronic conditions may perceive COVID-19 more severely than people who only have hypertension. The study also has a limitation in terms of residual confounding and long-term impact. I think this has to be reflected in the *Limitations* section.

Conflicts of Interest

None declared.

Reference

1. Bonner C, Cvejic E, Ayre J, Isautier J, Semsarian C, Nickel B, et al. The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case Control Study. JMIRx Med 2021 Mar 30;2(1):25610 [FREE Full text] [doi: 10.2196/25610]

Edited by E Meinert; submitted 11.03.21; this is a non–peer-reviewed article;accepted 11.03.21; published 30.03.21.

Please cite as:
Neupane D
Peer Review of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study"
JMIRx Med 2021;2(1):e28717
URL: https://xmed.jmir.org/2021/1/e28717
doi:10.2196/28717
PMID:

Peer Review Report

Peer Review of “Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis”

Anonymous Reviewer

Related Articles:

Companion article: https://preprints.jmir.org/preprint/21044

Companion article: https://med.jmirx.org/2021/1/e28742

Companion article: https://med.jmirx.org/2021/1/e21044/

(JMIRx Med 2021;2(1):e28679) doi:10.2196/28679

KEYWORDS

COVID-19; prediction; singular spectrum analysis; separability; eigenvalues; Saudi Arabia

This is a peer review submitted for the paper "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis".

Round 1 Review

Dear Author,

Thank you for the opportunity to review your paper [1].

I believe your manuscript would benefit from an editorial review prior to resubmission. This should include several elements: semantic and syntax review, native speaker edits, and formatting. Some of the words are illegible: modified presents as “modi ed”, different reads “di erent”, and many more. There are many words that appear incomplete or fragmented, which generally renders the manuscript illegible. This may have been due to a formatting bug during submission.

At this point, the paper should be reworked and then resubmitted so that an appropriate content review can take place.

Thank you and best.

Round 2 Review

Dear Author,

Thank you for addressing the formatting issue in your manuscript. However, I believe your manuscript would still benefit from an editorial review with respect to language prior to resubmission. This should include several elements: semantic review, syntax review, and native speaker edits (words such as “glop”). Also, I suggest removing references from the Abstract and adding them to the Introduction section of your paper.

Otherwise, I believe your paper will add an interesting viewpoint from a statistical perspective to COVID-19 modeling.

Thank you and best.

Reference

1. Alharbi N. Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis. JMIRx Med 2021 Mar 31;2(1):e21044 [FREE Full text] [doi: 10.2196/21044]

Edited by E Meinert; submitted 10.03.21; this is a non–peer-reviewed article;accepted 10.03.21; published 31.03.21.

Please cite as:
Anonymous Reviewer
Peer Review of "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis"
JMIRx Med 2021;2(1):e28679
URL: https://xmed.jmir.org/2021/1/e28679
doi:10.2196/28679
PMID:

Peer Review Report

Peer Review of “Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis”

Anonymous Reviewer

Related Articles:

Companion article: https://preprints.jmir.org/preprint/21044

Companion article: https://med.jmirx.org/2021/1/e28742

Companion article: https://med.jmirx.org/2021/1/e21044/

(JMIRx Med 2021;2(1):e28741) doi:10.2196/28741

KEYWORDS

COVID-19; prediction; singular spectrum analysis; separability; eigenvalues; Saudi Arabia

This is a peer review submitted for the paper "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis".

Round 1 Review

The paper is very interesting and very timely. I would suggest publishing it after minor revision.

Some information on the singular spectrum analysis technique could be added. Some additional information on COVID-19 cases could also be provided.

When one model outperforms another, it should be statistically tested. Considering the explanation of various parts as well as forecasting error, it is advisable to use a test that does not depend on the normality of error as well as h-step ahead forecasting.

Reference

1. Alharbi N. “Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis. JMIRx Med 2021 Mar 31;2(1):e21044 [FREE Full text] [doi: 10.2196/21044]

Edited by E Meinert; submitted 12.03.21; this is a non–peer-reviewed article;accepted 12.03.21; published 31.03.21.

Please cite as:
Anonymous Reviewer
Peer Review of “Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis”
JMIRx Med 2021;2(1):e28741
URL: https://xmed.jmir.org/2021/1/e28741
doi:10.2196/28741
PMID:

Authors' Response to Peer Reviews

Authors' Response to Peer Reviews of "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis"

Karim I Budhwani[1,2], PhD, DLA; Henna Budhwani[2], MPH, PhD; Ben Podbielski[3], MS, MBA

[1]CerFlux, Inc, Birmingham, AL, United States

[2]University of Alabama at Birmingham, Birmingham, AL, United States

[3]Protective Life Corp, Birmingham, AL, United States

Corresponding Author:
Karim I Budhwani, PhD, DLA
CerFlux Inc
2140 11th Avenue South
Suite 308
Birmingham, AL, 35205
United States
Phone: 1 8555000040
Fax: 1 8555000041
Email: ironman@cerflux.com

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22195

Companion article: https://med.jmirx.org/2021/1/e27103/

Companion article: https://med.jmirx.org/2021/1/e27257/

Companion article: https://med.jmirx.org/2021/1/e22195/

(JMIRx Med 2021;2(1):e27258) doi:10.2196/27258

KEYWORDS

infectious diseases; testing; per capita; population density; policy; coronavirus; SARS-CoV-2; COVID-19

Authors' response to peer reviews for "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis."

Response to Round 1 Reviews

Reviewer: Anonymous

General Comments

Dear anonymous reviewer [1], we would like to begin by conveying to you our deep appreciation for your assistance in refining this short paper [2] so that it is suitable for broader consumption. It is our aspiration that this paper will contribute positively to advancing knowledge in this domain. We have fully addressed all your recommendations and are pleased to submit a revised manuscript. Thank you for your expert assistance in this endeavor.

Specific Comments

Major Comments

You raise excellent points. We are happy to note that some of these points are a result of automatically transferring our manuscript from the preprint server. We submitted our manuscript originally to a preprint server with the goal of sharing our analysis and viewpoint in a timely and nonintimidating manner by way of a short report. The title, format, and manuscript text were rapidly copied from the general preprint server edition during the automatic transfer process.

The revised manuscript addresses the following:

1. The title has been updated to "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis."
2. Absolute terms from the preprint report have been modified.

3. The elevator vs football field "visual" expression was included deliberately in the original report as a means to make the role of density in SARS-CoV-2 viral transmission readily apparent to a broad audience. In order to address your concern, we have removed a reference to this expression in the *Results* section; however, in keeping with the original intent of reaching a broader audience, we would prefer to retain the expression in the *Introduction*.
4. We have included statements on limitations. Thank you for noting this gap.
5. We agree that a cost-effectiveness analysis is warranted after feasibility and acceptability have been established, but due in part to the word limit for short papers, we are unable to explore these differences. We believe that a paper on the costs and financial consequences of different testing strategies is warranted, potentially in follow-up analyses. Thank you for this recommendation.
6. In response to whether or not adjusted testing strategies based on population density (or similar measures) have been successfully done elsewhere: population density–based testing is novel, having (to our knowledge) only been employed in HIV research through network tracing in urban metropolitan areas. This gap in knowledge in terms of the benefit of population density testing is likely because we have not encountered many agents that are as infectious and persistent as SARS-CoV-2. This short paper is an initial step to illustrate to the scientific community that targeted approaches may be warranted when community spread occurs through close contact that is more likely in tightly packed communities.

Reviewer: AAA

General Comments

Dear reviewer AAA [3], we would like to begin by conveying to you our deep appreciation for your assistance in refining this short paper so that it is suitable for broader consumption. It is our aspiration that this paper will contribute positively to advancing knowledge in this domain. We have fully addressed all your recommendations and are pleased to submit a revised manuscript. Thank you for your expert assistance in this endeavor.

Specific Comments

Minor Comments

1. Your recommended heading change has been made in the revised manuscript.

Conflicts of Interest

None declared.

References

1. Anonymous. Peer Review of "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis". JMIRx Med 2021;2(1):e27103. [doi: 10.2196/27103]
2. Budhwani KI, Budhwani H, Podbielski B. Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis. JMIRx Med 2021;2(1):e22195. [doi: 10.2196/22195]
3. Allam AA. Peer Review of "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis". JMIRx Med 2021;2(1):e27257. [doi: 10.2196/27257]

Edited by E Meinert; submitted 18.01.21; this is a non–peer-reviewed article;accepted 19.01.21; published 03.02.21.

Please cite as:
Budhwani KI, Budhwani H, Podbielski B
Authors' Response to Peer Reviews of "Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis"
JMIRx Med 2021;2(1):e27258
URL: https://med.jmirx.org/2021/1/e27258/
doi:10.2196/27258
PMID:

Authors' Response to Peer Reviews

Authors' Response to Peer Reviews of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study"

Eduardo Atem De Carvalho[1*], PhD, MSc, BSc; Rogerio Atem De Carvalho[2*], DSC, MSc, MBA, BSc

[1]Center for Science and Technology, Universidade Estadual do Norte Fluminense, Campos, Brazil

[2]Innovation Hub, Instituto Federal Fluminense, Campos, Brazil

[*]all authors contributed equally

Corresponding Author:
Rogerio Atem De Carvalho, DSC, MSc, MBA, BSc
Innovation Hub
Instituto Federal Fluminense
R Cel Walter Kramer, 357
Campos, 28080-565
Brazil
Phone: 55 22 2737 5692
Email: ratembr@gmail.com

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22617

Companion article: https://med.jmirx.org/2021/1/e27536

Companion article: https://med.jmirx.org/2021/1/e27260/

Companion article: https://med.jmirx.org/2021/1/e22617

(JMIRx Med 2021;2(1):e27537) doi:10.2196/27537

KEYWORDS

COVID-19; pandemics; infection control; models; experimental; longitudinal studies; statistical modeling; epidemiology

Authors' response to peer reviews for "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study".

Response to Round 1 Reviews:

Dear Editor, we want to thank you for the opportunity offered by this prestigious journal.

Reviewer H:

General Comments

We appreciate the comments on our paper [1] from Reviewer H [2], which we address point by point. Regarding the size of the paper, we selected the most significant cases and the results obtained for Brazil, reducing the length in general. Regarding the wording, the paper underwent an initial edit after the first review round and was completely reviewed after its final acceptance. In order to highlight its scientific contribution, a bibliographical review on similar and recent articles was carried out, pointing out the need to provide a simple and at the same time effective model for analyzing epidemic curves. Additionally, we reorganized the paper to make its methodological section clearer and added statistical tests that provide support to our claims.

Specific Comments

Major Comments

1. As explained before, we focused on the most significant cases and removed the items cited by the reviewer as excessive.
2. We cited in the review an article that shows how often the Gaussian models failed in the predictions [3]. That section was removed, contributing to the reduction in the length of the paper.
3. As answered in the previous item, other authors already show that classic models have failed in their predictions. What we are looking for is a simple model that can be used by health authorities, and at the same time be

computationally efficient. In the text, using the appropriate reference [4], we show that the triangular distribution can replace the Gaussian distribution and its derivatives. In addition, we carried out the appropriate Kolmogorov-Smirnov tests, which prove this hypothesis mathematically for the cases studied. Therefore, we believe that we are contributing to the expansion of knowledge by including a frequency distribution that is still relatively unused in the field.

4. Citations have been properly formatted and organized.

Reviewer X:

General Comments

The authors are grateful for the comments of Reviewer X [5], which were all addressed point by point below. Regarding the structure of the paper, it now follows a format that includes a structured abstract, introduction, methods, results, and conclusions, highlighting its scientific aspects. Additionally, a brief bibliographic review was included that contributes to the justification of the paper's scientific contribution; statistical tests to assess the quality of the proposed model were also provided. In relation to the correct use of the terms of the field, the authors conducted a review and corrected them accordingly.

Specific Comments

1. The aim of the study is now clearly stated in the first two paragraphs of the Introduction section. A hypothesis was successfully tested using the Kolmogorov-Smirnov method in the Results section.
2. The subtopic "Nondimensional Characteristics of Epidemic Cycles" introduces the triangular distribution, its assumptions, and its limitations in the context of the COVID-19 pandemic. It is complemented by the subtopic "Predictability by Similarity."
3. In the Methods section, the source and date of data collection are explicitly stated (Worldometer's COVID-19 portal, as of July 9, 2020). A copy of this data has now been provided in a separate Excel spreadsheet.
4. The terms were checked and fixed; to the best of our knowledge, they now match those used by experts in the field.

Response to Round 2 Reviews:

Regarding the journal instructions: "Using the structure used in this paper, please consolidate manuscript 23997 and 23998 into one cohesive narrative, taking into account peer-review feedback provided by the reviewers on those submissions. This way, we can present one paper with your aggregated findings in JMIRx." We followed these recommendations and integrated the three articles originally called "Identification of Patterns in Epidemic Cycles and Methods for Estimating Their Duration: COVID-19 Case Study," "COVID-19: Time-Dependent Effective Reproduction Number and Sub-notification Effect Estimation Modeling," and "COVID-19: Estimation of the Actual Onset of Local Epidemic Cycles, Determination of Total Number of Infective, and Duration of the Incubation Period" into a single narrative. The resulting paper represents not only the combination of the content of the three others but an integrated narrative that describes the statistical framework developed by us to analyze the epidemic cycles. Thus, given that the resulting content reflects this integrated work, we find it more coherent to change the title of the article to "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study." In order to make the resulting text more fluid, we concentrated on the data analyzed in three countries (Germany, Italy, and Sweden), leaving the case studies related to Brazil to Multimedia Appendix 3. Other cases originally studied during the development of the statistical framework can still be found in the preprints duly referenced in the text. In addition, all the data obtained from the referred public databases, as well as all the calculations carried out both in the main paper and in Multimedia Appendix 3, are organized in three different worksheets (Multimedia Appendices 2-4), in order to facilitate the verification and reproduction of the results by readers.

References

1. De Carvalho EA, De Carvalho RA. A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study. JMIRx Med 2021 Mar 18;2(1):e22617 [FREE Full text] [doi: 10.2196/22617]
2. Facente SN. Peer Review of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study". JMIRx Med 2021 Mar 18;2(1):e27536. [doi: 10.2196/27536]
3. Park SW, Bolker BM, Champredon D, Earn DJD, Li M, Weitz JS, et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. J R Soc Interface 2020 Jul;17(168):20200144 [FREE Full text] [doi: 10.1098/rsif.2020.0144] [Medline: 32693748]
4. Kotz S, René van Dorp J. Beyond Beta - Other Continuous Families of Distributions with Bounded Support and Applications. Singapore: World Scientific Publishing Company; 2004.
5. Salman M. Peer Review of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study". JMIRx Med 2021 Mar 18;2(1):e27260. [doi: 10.2196/27260]

Edited by E Meinert; submitted 27.01.21; this is a non–peer-reviewed article;accepted 27.01.21; published 18.03.21.

Please cite as:
De Carvalho EA, De Carvalho RA
Authors' Response to Peer Reviews of "A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study"
JMIRx Med 2021;2(1):e27537
URL: https://xmed.jmir.org/2021/1/e27537
doi:10.2196/27537
PMID:

Authors' Response to Peer Reviews

Authors' Response to Peer Reviews of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study"

Carissa Bonner[1], PhD; Erin Cvejic[1], PhD; Julie Ayre[1], PhD; Jennifer Isautier[1], MSc; Christopher Semsarian[2,3,4], PhD; Brooke Nickel[1], PhD; Carys Batcup[1], MSc; Kristen Pickles[1], PhD; Rachael Dodd[1], PhD; Samuel Cornell[1], MSc; Tessa Copp[1], PhD; Kirsten J McCaffery[1], PhD

[1]Sydney Health Literacy Lab, School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
[2]Agnes Ginges Centre for Molecular Cardiology, Centenary Institute, The University of Sydney, Sydney, Australia
[3]Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
[4]Department of Cardiology, Royal Prince Alfred Hospital, Sydney, Australia

Corresponding Author:
Carissa Bonner, PhD
Sydney Health Literacy Lab, School of Public Health
Faculty of Medicine and Health
The University of Sydney
Edward Ford Building A27, Rm 128A
Sydney, NSW 2006
Australia
Phone: 61 2 9351 7125
Email: carissa.bonner@sydney.edu.au

Related Articles:

Companion article: https://preprints.jmir.org/preprint/25610

Companion article: https://med.jmirx.org/2021/1/e28714/

Companion article: https://med.jmirx.org/2021/1/e28717/

Companion article: https://med.jmirx.org/2021/1/e25610/

(JMIRx Med 2021;2(1):e28718) doi:10.2196/28718

KEYWORDS

public health; global health; COVID-19; hypertension; risk; strategy; mental health; behavior; response; anxiety; vaccine; retrospective; perception; prevention; intention

Authors' response to peer reviews for "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study."

Response to Reviews

The authors of the manuscript [1] are grateful to the editor and reviewers [2,3] for their invaluable input and feedback.

Reviewer G

Round 1

Specific Comments: Major

Thank you for your review [2]. We wrote this as a brief correspondence piece for rapid publication as a preprint because COVID-19 research and public health communications were rapidly evolving at this time. However, we have now rewritten the paper in standard paper format to address your concerns—this includes a much more detailed introduction and rationale, more explanations in the *Methods* section including assumptions, and an expanded discussion. Please see our detailed responses below.

Specific Comments: Minor

We have addressed the issues highlighted in the rewritten manuscript. Please see our detailed responses below.

Detailed Responses

Title/Abstract and References

1. We have changed the title to better reflect our methodology.

2. We have rewritten our work as a full paper rather than a short correspondence, including suggested references and other research that has emerged since the publication of our rapid preprint. The media references are important as they provide context for the study during a rapidly changing COVID-19 response.

Introduction

1 and 2. We have included the references/points mentioned by the reviewer and a summary of other research that has emerged since the publication of our rapid preprint. Please note the *New England Journal of Medicine* paper has been retracted so we have not included this.

Methods

1. We have included more details on methods to address the reviewer's points. This includes new sections on the setting, matching, and analysis.

2. We confirm that the original ethics approval covers all subsequent surveys and amendments. We have clarified this in the manuscript.

3. Another preprint from our study [4] that describes the sample at different time point shows it is comparable between April and June. We acknowledge that the respondents who remained in the study were likely more motivated and interested in COVID-19 prevention than those who dropped out. This is mentioned in the *Discussion* section. However, since the study design is not a randomized controlled trial, cases were matched to controls at the same time point, with demographic characteristics controlled for in the analyses, so our key comparison findings should not be affected by those differences.

We have also now performed exploratory analyses of the hypertension subsample by whether they were invited and returned for follow-up, compared to those who were not invited or did not return. We have included a text summary in the paper: "Those who were invited and returned for follow-up were similar for age and gender but had higher levels of education (P=.02) and were more likely to have adequate health literacy (P=.009)."

4. We have added more details about measures. This followed a US study published in the *Annals of Internal Medicine* [5], with whose authors we are collaborating on an international comparison.

"Participants were asked if they had any of the following conditions: asthma, chronic obstructive pulmonary disease, high blood pressure (hypertension), cancer, heart disease, stroke, diabetes, depression, anxiety."

5. We have added more details in the new *Methods* section.

6. We have added more details in the new *Methods* section.

7. (a) The estimates referred to as "MMDs" from the linear models are marginal mean differences, not maximum mean discrepancy. This abbreviation is noted in the first row of Table 2. We have now added the abbreviations to the footnote of the table to avoid confusion and clarified the first use of MMD when reporting results in text.

(b) In regard to the use of a modified Poisson approach and reporting of relative risks: with increasing event rates, the difference between an odds ratio (as estimated by logistic regression) and the risk ratio (as estimated from a log-binomial or modified Poisson model) also increases, with odds ratios often incorrectly interpreted as if they are risks. As the reviewer points out, the study design is cross-sectional, so a risk/prevalence ratio is typically considered more appropriate and conceptually easier to interpret than an odds ratio. Although log-binomial regression can also be used to estimate the risk ratio, it is often criticized for producing confidence intervals that are narrower than they should be (ie, due to smaller than expected standard errors) and may also fail to converge. For this reason, we have employed a modified Poisson approach [6], which generates coefficients that, when exponentiated, represent the risk ratio, with corresponding confidence intervals of an appropriate width. As for reporting relative risk, numerically, the risk ratio/relative risk and prevalence ratio are identical, differing only in their interpretation based on the study design. In line with the reviewer's comment, we have changed the language used and describe the effect as an adjusted prevalence ratio rather than adjusted relative risk to better reflect the study design.

8. We have added a statement on this.

Results

1. We have added the footnote. Pairwise comparisons showed no statistically significant differences in age, gender, education, or health literacy between the hypertension and control groups (see the section on matching). We have explained social distancing and patient activation, and clarified the prescription question in the *Methods* section.

2. We have added this.

3. Please see our response above regarding MMD (marginal mean differences from linear regression models). As for the social distancing score, this is a typographical error and has been corrected.

4. We have clarified that these are two separate questions.

5. We have moved this to the *Discussion* section with additional explanations.

6. We have clarified this.

Discussion and Conclusions

1-6. We have rewritten the article as a full paper rather than a short correspondence, including a more expansive discussion to address the points mentioned. We have highlighted key findings upfront, discussed different perspectives including access to care, clarified that we only measured vaccination intentions throughout, discussed the implications of the

limitations, included points about misinformation on social media, and have more carefully explained our conclusions.

Reviewer AM

Round 1

Thank you for your review [3]. We have addressed the comments as follows.

Title

We have revised the title to better reflect the study methods.

Abstract

We have revised this as suggested.

Methods

1. We have added this.

2. We have expanded the *Methods* section to clarify this.

Results

1. Our controls were defined as not having comorbidities thought to be relevant to COVID-19 outcomes at the time of the study. Other medications (eg, contraception or unrelated conditions) could have been taken, but we did not ask for these details in this survey.

2. Per our previous responses, we have now expanded our *Methods* and *Statistical Analysis* sections, clarifying that the MMD was calculated for continuous outcomes using linear regression models.

Although the risk perception measure has a restricted range of 0-10, the normality assumption of a general linear model relates to the distribution of the residuals (which should also be homogenous across the fitted values). These assumptions were explored graphically (via a histogram of residuals with superimposed normal density and a plot of the residuals against the fitted values with a superimposed smoothed lowess line), and was deemed to be sufficiently satisfied.

Notably, an alternative analysis approach for ordinal Likert-scale data would be to apply an ordinal logistic regression model. We have explored this option given the reviewer's comment; this elicited comparable outcomes. However, there was substantially more difficulty associated with the interpretation given the outcome of such a model is the adjusted odds ratio of responding one unit higher on the response scale for cases relative to controls. As such, we feel the application of linear regression remains a more suitable option for this outcome variable.

3. We have added this.

Discussion/Limitations

We have included a more expansive introduction and discussion to address the reviewer's points. Our COVID-19 risk perception and vaccination intention scores were very high across groups, indicating a possible ceiling effect, but this is consistent with other Australian surveys, which is explained in our discussion. We have explained why we isolated the effect of hypertension from other comorbidities in the expanded paper with more details in the *Introduction* and *Methods* sections.

Round 2

We have added the ethics review/approval number in this version of the manuscript as requested.

References

1. Bonner C, Cvejic E, Ayre J, Isautier J, Semsarian C, Nickel B, et al. The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case Control Study. JMIRx Med 2021 Mar 30;2(1):e25610 [FREE Full text] [doi: 10.2196/25610]
2. Delgado-Ron JA. Peer Review of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study". JMIRx Med 2021 Mar 30;2(1):e28714 [FREE Full text] [doi: 10.2196/28714]
3. Neupane D. Peer Review of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study". JMIRx Med 2021 Mar 30;2(1):e28717 [FREE Full text] [doi: 10.2196/28717]
4. Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Ayre J, et al. COVID-19: Beliefs in misinformation in the Australian community. medRxiv. Preprint posted online August 06, 2020. [FREE Full text] [doi: 10.1101/2020.08.04.20168583]
5. Wolf MS, Serper M, Opsasnick L, O'Conor RM, Curtis L, Benavente JY, et al. Awareness, Attitudes, and Actions Related to COVID-19 Among Adults With Chronic Conditions at the Onset of the U.S. Outbreak: A Cross-sectional Survey. Ann Intern Med 2020 Jul 21;173(2):100-109 [FREE Full text] [doi: 10.7326/M20-1239] [Medline: 32271861]
6. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004 Apr 01;159(7):702-706. [doi: 10.1093/aje/kwh090] [Medline: 15033648]

Edited by E Meinert; submitted 11.03.21; this is a non–peer-reviewed article;accepted 11.03.21; published 30.03.21.

Please cite as:
Bonner C, Cvejic E, Ayre J, Isautier J, Semsarian C, Nickel B, Batcup C, Pickles K, Dodd R, Cornell S, Copp T, McCaffery KJ Authors' Response to Peer Reviews of "The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study"
JMIRx Med 2021;2(1):e28718
URL: https://xmed.jmir.org/2021/1/e28718
doi:10.2196/28718
PMID:

Authors' Response to Peer Reviews

Author's Responses to Peer Reviews of "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis"

Nader Alharbi[1], PhD

King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Corresponding Author:
Nader Alharbi, PhD
King Saud bin Abdulaziz University for Health Sciences
King Abdullah International Medical Research Center
Prince Mutib Ibn Abdullah Ibn Abdulaziz Rd, Ar Rimayah
Riyadh
Saudi Arabia
Phone: 966 114299999 ext 95590
Email: alharbina@ksau-hs.edu.sa

Related Articles:

Companion article: https://preprints.jmir.org/preprint/21044

Companion article: https://med.jmirx.org/2021/1/e28679/

Companion article: https://med.jmirx.org/2021/1/e28741/

Companion article: https://med.jmirx.org/2021/1/e21044/

This is a corrected version. See correction statement: https://xmed.jmir.org/2021/2/e29879

(JMIRx Med 2021;2(1):e28742) doi:10.2196/28742

KEYWORDS

COVID-19; prediction; singular spectrum analysis; separability; eigenvalues; Saudi Arabia

These are the author's response to peer reviews for "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis".

Response to Round 1 Reviews

First, I wish to thank the editor and all the reviewers for their enlightening comments and observations, which I strongly believe have increased the quality of this manuscript [1]. I am indeed grateful for the time they have devoted to this paper. I hope you will be pleased to learn that I have taken on board and addressed all observations and comments.

Anonymous [2]

Specific Comments

1. Thank you for this comment [2]. I have revised the discussion of the singular spectrum analysis technique and COVID-19 cases according to the referee's suggestion, and I added more information and citations.
2. I appreciate the reviewer's insightful comment. The test was used, and the results are provided. The recommended paper and another related paper were cited.

Anonymous [3]

General Comments

Thank you for pointing out the formatting issues to me [3]. I agree, they must have been due to a formatting bug during submission. I hope now you will be pleased, as I have addressed all observations and submitted the paper as a Word file.

Response to Round 2 Review

I appreciate the reviewer's insightful comments and suggestions.

1. The whole manuscript has been reviewed and edited by a native speaker as suggested.

2. Moreover, the references have been removed from the Abstract and added to the Introduction section.

References

1. Alharbi N. Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis. JMIRx Med 2021 Mar 31;2(1):e21044 [FREE Full text] [doi: 10.2196/21044]
2. Anonymous Reviewer. Peer review of "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis". JMIRx Med 2021 Mar 31;2(1):e28741 [FREE Full text] [doi: 10.2196/28741]
3. Anonymous Reviewer. Peer review of "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis". JMIRx Med 2021 Mar 31;2(1):e28679 [FREE Full text] [doi: 10.2196/28679]

Edited by E Meinert; submitted 12.03.21; this is a non–peer-reviewed article;accepted 12.03.21; published 31.03.21.

Please cite as:
Alharbi N
Author's Responses to Peer Reviews of "Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis"
JMIRx Med 2021;2(1):e28742
URL: https://xmed.jmir.org/2021/1/e28742
doi:10.2196/28742
PMID:

Short Paper

Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis

Karim I Budhwani[1,2*], PhD, DLA; Henna Budhwani[2*], MPH, PhD; Ben Podbielski[3*], MS, MBA

[1]CerFlux Inc, Birmingham, AL, United States

[2]University of Alabama at Birmingham, Birmingham, AL, United States

[3]Protective Life Corp, Birmingham, AL, United States

[*]all authors contributed equally

Corresponding Author:
Karim I Budhwani, PhD, DLA
CerFlux Inc
2140 11th Avenue South
Suite 308
Birmingham, AL, 35205
United States
Phone: 1 8555000040
Fax: 1 8555000041
Email: ironman@cerflux.com

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22195

Companion article: https://med.jmirx.org/2021/1/e27103/

Companion article: https://med.jmirx.org/2021/1/e27257/

Companion article: https://med.jmirx.org/2021/1/e27258/

Abstract

Background: SARS-CoV-2 transmission risk generally increases with the proximity of those shedding the virus to those susceptible to infection. Thus, this risk is a function of both the number of people and the area they occupy. However, the latter continues to evade the COVID-19 testing policy.

Objective: The aim of this study is to analyze per capita COVID-19 testing data reported for Alabama to evaluate whether testing realignment along population density, rather than density agnostic per capita, would be more effective.

Methods: Descriptive statistical analyses were performed for population, density, COVID-19 tests administered, and positive cases for all 67 Alabama counties.

Results: Tests reported per capita appeared to suggest widespread statewide testing. However, there was little correlation (r=0.28, P=.02) between tests per capita and the number of cases. In terms of population density, new cases were higher in areas with a higher population density, despite relatively lower test rates as a function of density.

Conclusions: Increased testing in areas with lower population density has the potential to induce a false sense of security even as cases continue to rise sharply overall.

(JMIRx Med 2021;2(1):e22195) doi:10.2196/22195

KEYWORDS

infectious diseases; testing; per capita; population density; policy; coronavirus; SARS-CoV-2; COVID-19

Introduction

COVID-19 testing is typically measured per capita; tests and cases are reported per million globally while local authorities report counts per 100,000 people [1-3]. This approach is simple and generally well accepted both in economic spheres and in health care research. However, this simplicity may shroud an underlying fallacy in applying per capita models to test the transmission characteristics of SARS-CoV-2. The transmission risk profile for 20 people in an elevator is substantially different from that of 20 people spread across a football field; this was the fundamental premise for social distancing and lockdowns to "flatten the curve." Moreover, population density can impede [4] implementation of protective distancing measures. Population density has also been implicated [5] in COVID-19 mortality. In this two-part study, we analyze per capita COVID-19 testing data reported for Alabama to evaluate whether testing realignment along population density, rather than density agnostic per capita, would be more effective, as Alabama is one of several states currently experiencing notable increases in new cases.

Methods

Population characteristics and population density for all 67 Alabama counties were obtained from the 2018 American Community Survey (US Census Bureau). The number of tests administered and positive cases of COVID-19 are updated daily by the Alabama Department of Public Health. These data were obtained on May 18, 2020, for initial assessment and again on June 15, 2020, for prospective analysis. Descriptive statistical analyses were performed to calculate the total number of tests per 100,000 people using the county population as the denominator, and subsequently dividing this by county population density, density squared, and square root of density as illustrative proxies [6,7] of more complex population density test rate models. All study data were publicly available, thereby obviating institutional review board approval.

Results

The first heatmap presented in Figure 1 appears to indicate widespread testing per 100,000 people [8] by county. However, this heatmap does not distinguish sparsely populated areas that could inherently provide spatial distancing from those that are densely populated (Figure 1B) [9]. Overlaying the two (Figure 1C) provides a sense of magnitude by which we may be overtesting in areas with a natural spatial defense against transmission while severely undertesting in areas with an elevated risk of transmission.

In the second part of the study, conducted during the phased economic re-engagement, data were collected to prospectively analyze the distribution of tests and cases vis-à-vis population density. Tests reported per 100,000 during this period, once again, appeared to indicate widespread statewide testing. However, there was little correlation (r=0.28, P=.02) between tests per capita and the number of cases. As anticipated [10], new cases were disproportionately more prevalent in densely populated areas (Figure 2), despite relatively fewer tests per population density, suggesting that cases in these areas may be understated.

Figure 1. Per capita and population density heatmaps for COVID-19 tests between April 1 and May 18, 2020. (A) Heatmap of tests per 100,000. (B) Population density heatmap distinguishing sparsely populated areas from those that are densely populated. (C) Overlaying the two shows current testing by population density. Without a population density–driven testing approach, the risk of deriving a false sense of security is greater.

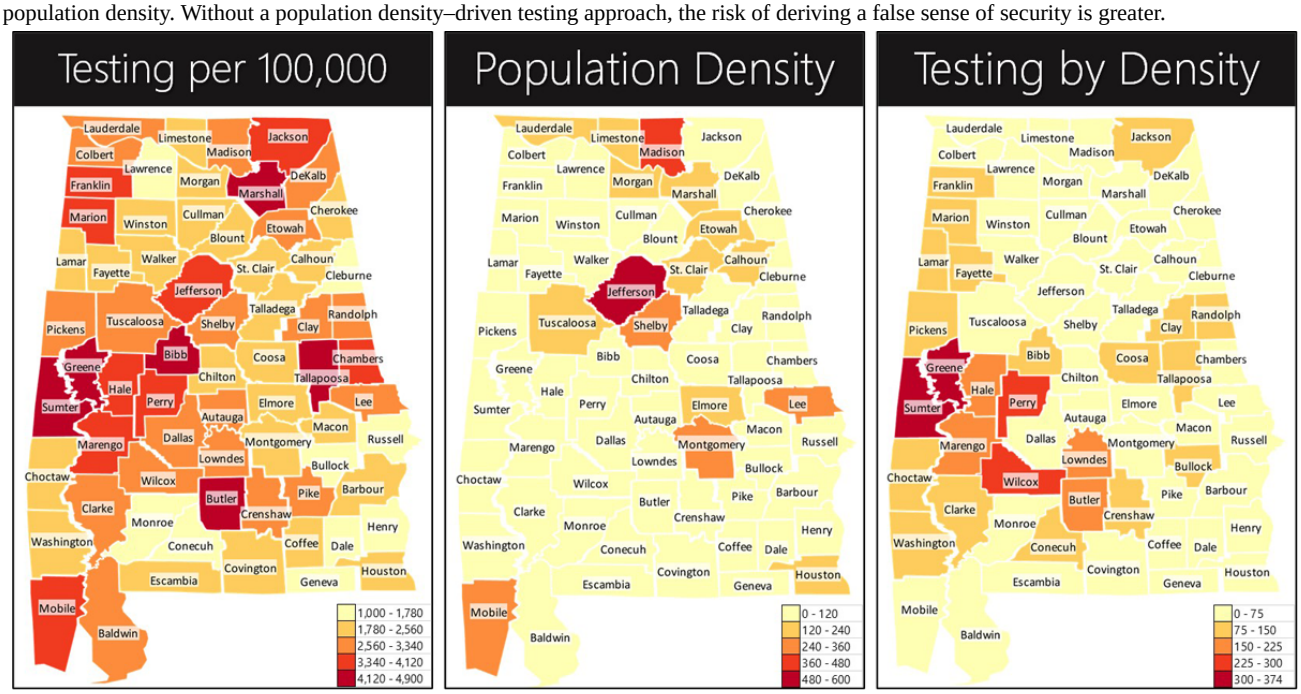
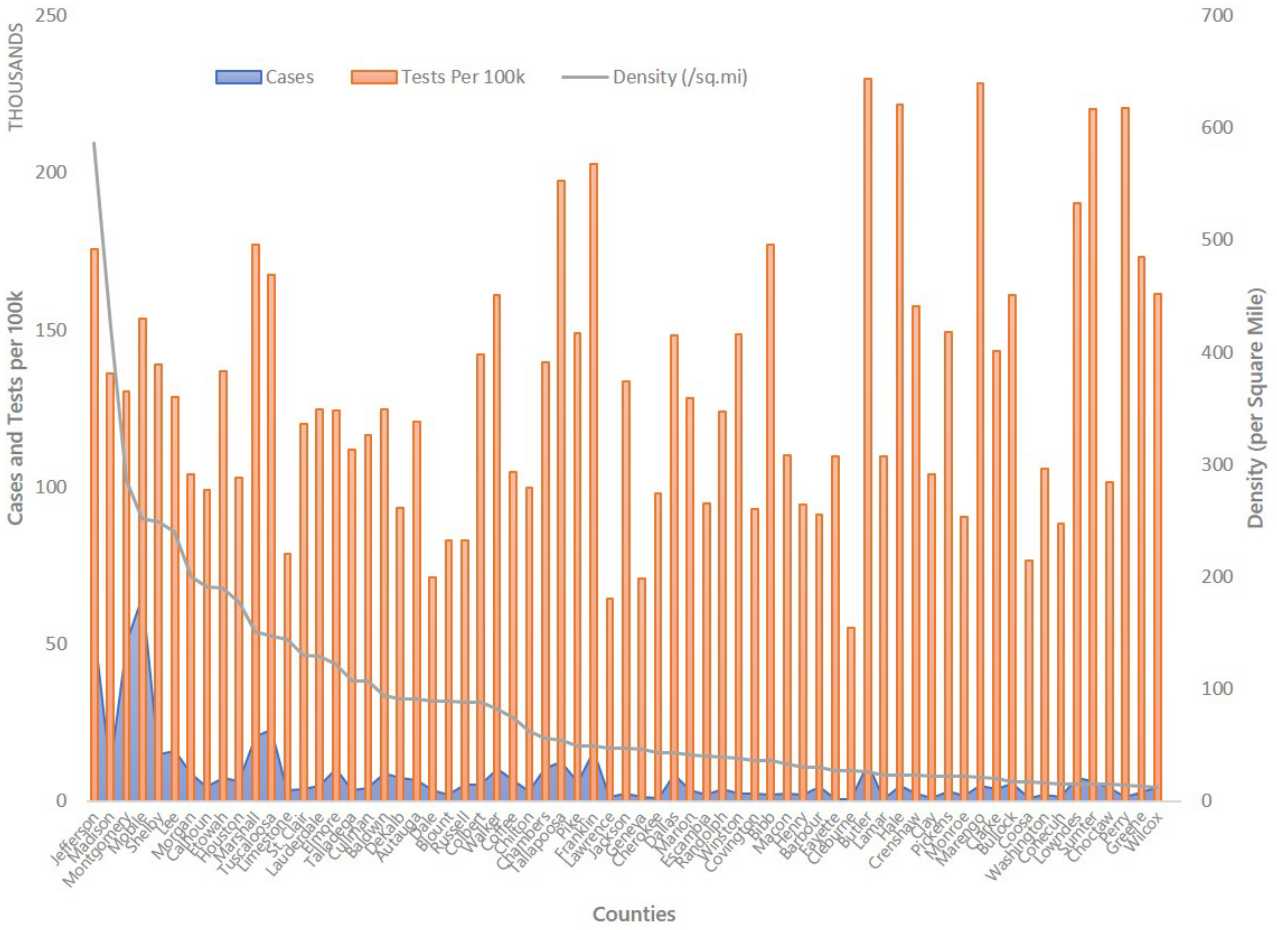


Figure 2. COVID-19 testing during the phased reopening of the Alabama economy from May 18 to June 15, 2020. Tests reported per 100,000 during this period also appeared to indicate widespread statewide testing. However, there was little correlation (r=0.28, P=.02) between tests per capita and the number of cases. In terms of population density, new cases were higher in areas with higher population density, despite relatively lower test rates as a function of density. This suggests that a population density–driven testing strategy would not only allow for more effective allocation but could also reduce the risk of understating cases in areas with high population density.



Discussion

The current standard of population density agnostic per capita reporting could induce a sense of false security while simultaneously accelerating infection in economic nerve centers. The contrast among the heatmaps, as well as subsequent prospective analysis of tests and cases, unveil the scale of testing disparity. A robust testing strategy would presumably figure prominently in the calculus for any phased reopening of economies and associated near-term paths to societal normalcy and economic recovery. Consequently, disparities in testing induced by a density agnostic testing approach could undermine balancing measures aimed at saving lives and livelihoods, thereby leading to a prolonged recession, or dare we say, a depression [11,12].

Although we use Alabama for illustration, most states report statistics in this manner, making our processes replicable in other states. This said, limitations of our approach should be considered when extending findings. Namely, population density–driven testing has not be extensively evaluated for feasibility and acceptability, and, during this pandemic, gaps in public health monitoring and surveillance data [5], particularly from rural communities, have emerged, leading to concerns related to data reliability.

On a positive note, resolving this is not intractable. Heatmaps of retail and payroll activity are unsurprisingly similar to population density. This is where the innate intertwining of public health and economic well-being around the “location, location, location” axis can be synergistic. For instance, by adjusting the distribution of testing capacity to also account for population density, we could improve monitoring and response to blunt the speed and spread of the virus while also safeguarding both retail activity and economic nerve centers across the country.

Authors' Contributions

All authors contributed to the writing of this manuscript and have approved the final version.

Conflicts of Interest

None declared.

References

1. United States COVID-19 Cases and Deaths by State. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/covid-data-tracker/ [accessed 2020-01-05]
2. Lyu W, Wehby GL. Comparison of Estimated Rates of Coronavirus Disease 2019 (COVID-19) in Border Counties in Iowa Without a Stay-at-Home Order and Border Counties in Illinois With a Stay-at-Home Order. JAMA Netw Open 2020 May 01;3(5):e2011102 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.11102] [Medline: 32413112]
3. Memish ZA, Aljerian N, Ebrahim SH. Tale of three seeding patterns of SARS-CoV-2 in Saudi Arabia. The Lancet Infectious Diseases 2021 Jan;21(1):26-27. [doi: 10.1016/s1473-3099(20)30425-4]
4. Gibson L, Rush D. Novel Coronavirus in Cape Town Informal Settlements: Feasibility of Using Informal Dwelling Outlines to Identify High Risk Areas for COVID-19 Transmission From A Social Distancing Perspective. JMIR Public Health Surveill 2020 Apr 06;6(2):e18844 [FREE Full text] [doi: 10.2196/18844] [Medline: 32250283]
5. Bray I, Gibson A, White J. Coronavirus disease 2019 mortality: a multivariate ecological analysis in relation to ethnicity, population density, obesity, deprivation and pollution. Public Health 2020 Aug;185:261-263 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.056] [Medline: 32693249]
6. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. Math Biosci 2013 Aug;244(2):125-134 [FREE Full text] [doi: 10.1016/j.mbs.2013.04.013] [Medline: 23665296]
7. Smith MJ, Telfer S, Kallio ER, Burthe S, Cook AR, Lambin X, et al. Host-pathogen time series data in wildlife support a transmission function between density and frequency dependence. Proc Natl Acad Sci U S A 2009 May 12;106(19):7905-7909 [FREE Full text] [doi: 10.1073/pnas.0809145106] [Medline: 19416827]
8. COVID-19 in Alabama. Alabama Department of Public Health. 2020. https://dph1.adph.state.al.us/covid-19/ [accessed 2020-01-05].
9. US Census Data. US Census Bureau. 2018. URL: https://data.census.gov/cedsci/ [accessed 2020-05-01]
10. Rubin D, Huang J, Fisher BT, Gasparrini A, Tam V, Song L, et al. Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States. JAMA Netw Open 2020 Jul 01;3(7):e2016099 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16099] [Medline: 32701162]
11. Felsenthal M. COVID-19 to Plunge Global Economy into Worst Recession since World War II. The World Bank. 2020 Jun 8. URL: https://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-into-worst-recession-since-world-war-ii.print [accessed 2021-01-19]
12. Global Economic Prospects. World Bank Group. 2020 Jun. URL: https://openknowledge.worldbank.org/bitstream/handle/10986/33748/9781464815539.pdf [accessed 2021-01-19]

Edited by G Eysenbach, E Meinert; submitted 27.08.20; peer-reviewed by Anonymous, AA Allam; comments to author 04.11.20; revised version received 22.12.20; accepted 24.12.20; published 03.02.21.

Please cite as:
Budhwani KI, Budhwani H, Podbielski B
Evaluating Population Density as a Parameter for Optimizing COVID-19 Testing: Statistical Analysis
JMIRx Med 2021;2(1):e22195
URL: https://med.jmirx.org/2021/1/e22195/
doi:10.2196/22195
PMID:33725028

Original Paper

A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study

Eduardo Atem De Carvalho[1*], PhD, MSc, BSc; Rogerio Atem De Carvalho[2*], DSC, MSc, MBA, BSc

[1]Center for Science and Technology, Universidade Estadual do Norte Fluminense, Campos, Brazil

[2]Innovation Hub, Instituto Federal Fluminense, Campos, Brazil

[*]all authors contributed equally

Corresponding Author:
Rogerio Atem De Carvalho, DSC, MSc, MBA, BSc
Innovation Hub
Instituto Federal Fluminense
R Cel Walter Kramer, 357
Campos, 28080-565
Brazil
Phone: 55 22 27375692
Email: ratem@iff.edu.br

Related Articles:

Companion article: https://preprints.jmir.org/preprint/22617

Companion article: https://med.jmirx.org/2021/1/e27536

Companion article: https://med.jmirx.org/2021/1/e27260/

Companion article: https://med.jmirx.org/2021/1/e27537

Abstract

Background: Since the beginning of the COVID-19 pandemic, researchers and health authorities have sought to identify the different parameters that drive its local transmission cycles to make better decisions regarding prevention and control measures. Different modeling approaches have been proposed in an attempt to predict the behavior of these local cycles.

Objective: This paper presents a framework to characterize the different variables that drive the local, or epidemic, cycles of the COVID-19 pandemic, in order to provide a set of relatively simple, yet efficient, statistical tools to be used by local health authorities to support decision making.

Methods: Virtually closed cycles were compared to cycles in progress from different locations that present similar patterns in the figures that describe them. With the aim to compare populations of different sizes at different periods of time and locations, the cycles were normalized, allowing an analysis based on the core behavior of the numerical series. A model for the reproduction number was derived from the experimental data, and its performance was presented, including the effect of subnotification (ie, underreporting). A variation of the logistic model was used together with an innovative inventory model to calculate the actual number of infected persons, analyze the incubation period, and determine the actual onset of local epidemic cycles.

Results: The similarities among cycles were demonstrated. A pattern between the cycles studied, which took on a triangular shape, was identified and used to make predictions about the duration of future cycles. Analyses on effective reproduction number (R_t) and subnotification effects for Germany, Italy, and Sweden were presented to show the performance of the framework introduced here. After comparing data from the three countries, it was possible to determine the probable dates of the actual onset of the epidemic cycles for each country, the typical duration of the incubation period for the disease, and the total number of infected persons during each cycle. In general terms, a probable average incubation time of 5 days was found, and the method used here was able to estimate the end of the cycles up to 34 days in advance, while demonstrating that the impact of the subnotification level (ie, error) on the effective reproduction number was <5%.

Conclusions: It was demonstrated that, with relatively simple mathematical tools, it is possible to obtain a reliable understanding of the behavior of COVID-19 local epidemic cycles, by introducing an integrated framework for identifying cycle patterns and calculating the variables that drive it, namely: the R_t, the subnotification effects on estimations, the most probable actual cycles start dates, the total number of infected, and the most likely incubation period for SARS-CoV-2.

(JMIRx Med 2021;2(1):e22617) doi:10.2196/22617

KEYWORDS

COVID-19; SARS-CoV-2; pandemics; infection control; models; experimental; longitudinal studies; statistical modeling; epidemic cycles

Introduction

The analysis of the life cycles of any epidemic involves the analysis of a series of quantitative parameters that govern these cycles and which, given the inherent uncertainty of these events, are generally treated by statistical models. For a number of practical reasons, the registration of deaths and of infections are inevitably imprecise, although these numbers can be corrected over time. Therefore, with the COVID-19 pandemic, a subject that immediately became the center of debates and different studies was the characterization of the different local epidemic cycles and their corresponding variables. Local cycles are those that have occurred or occur in specific countries, regions, or cities, and not the pandemic cycle as a whole, as the virus does not spread instantly across continents. Thus, it can be seen that some countries were in more advanced epidemic stages than others whose first infections were detected later. In other words, as expected, different "infection windows" coexist in parallel in different locations, with some locations at a more advanced stage, while others present more "delayed" cycles. Thus, numerically analyzing the behavior of early cycles was the measure undertaken by a series of researchers.

Although it is not the only one, as will be seen in this paper, the reproduction number is considered the central variable in the analysis of epidemic cycles. In order to determine the reproduction number, different categories of models have been proposed: artificial neural networks [1], Poisson [2,3], exponential [4], Markov chain [5], Gaussian [6,7], Weibull [8], Logistic-S [9], and moving averages [10]. Most research tries to frame the local epidemic cycles into Gaussian and/or Weibull behaviors, creating complex models that still led to errors in predictions, as we now know. More importantly, Park et al [11] showed that the initial models, most based on the Gaussian distribution and its derivatives, failed to make their predictions. After observing these findings, we saw that there was room to propose a framework that would provide an efficient and more comprehensive analysis of the epidemic cycles, going beyond the calculation of the reproduction number. Moreover, it would be both easy to understand and to compute, since local authorities, especially in low-income countries, do not always have statistical experts at their disposal to propose, calibrate, and analyze the results of complex models. Thus, based on experimental and publicly available data, we produced a series of studies that initially dealt with the identification of patterns in epidemic cycles and their use for predicting deaths [12], time-dependent effective reproduction number (R_t) and subnotification effect estimation modeling [13], and finally, estimation of the actual onset of local epidemic cycles, determination of the total number of infected, and the duration of the incubation period [14]. In this paper, these findings are integrated and summarized in a coherent framework.

Methods

Based on experimental data, the framework proposed here is divided into four parts: (1) applying the moving averages method and identifying the parameters of the epidemic cycle patterns, which are used to predict the number of future deaths in local epidemics, (2) modeling the R_t and (3) the effects of subnotification, and (4) applying the logistic model associated to a novel inventory model to obtain the final count for the total infected, the daily infection rate and lag time, and the incubation period.

Patterns of Epidemic Cycles

Our method began with the observation of several cycles in western countries where the pandemic hit earlier, especially in Europe. From there, patterns were identified and predictions were applied. The attempt to describe the different epidemic cycles that make up the current pandemic often comes up against the quality of the data that is made public. Most data made public are based on "date of recording," which is different of "day of death," meaning that the date that a given set of deaths are recorded in the public health statistics systems is not necessarily the date they occurred on; given the usual bureaucratic procedures, recording may be delayed.

The fact is that the distribution of fatalities suffers a distortion that generates a "saw" appearance in graphs such that on weekends there is a clear absence of death records, followed by an explosion of values at the beginning of the week. A simple technique that softens this effect is to apply the so-called moving average method (MAM), in which the daily value of deaths is replaced by the sum of the previous 6 days with the current day, divided by 7; in other words, the average of the week ended in the current day. In particular, MAMI (MAM with initial value) will be used here, which entails assigning the average of the 7 days to the first day of the week (Sunday).

In the period in which the data were obtained and analyzed (first week of July 2020), several cities, regions, states, and countries had already completed what will be called here the most lethal cycle of the epidemic (MLCE), which is when the number of deaths increases daily, on average, until it reaches a peak and then begins to decrease continuously until it reaches a minimum value. After this period, the occurrence of deaths continues

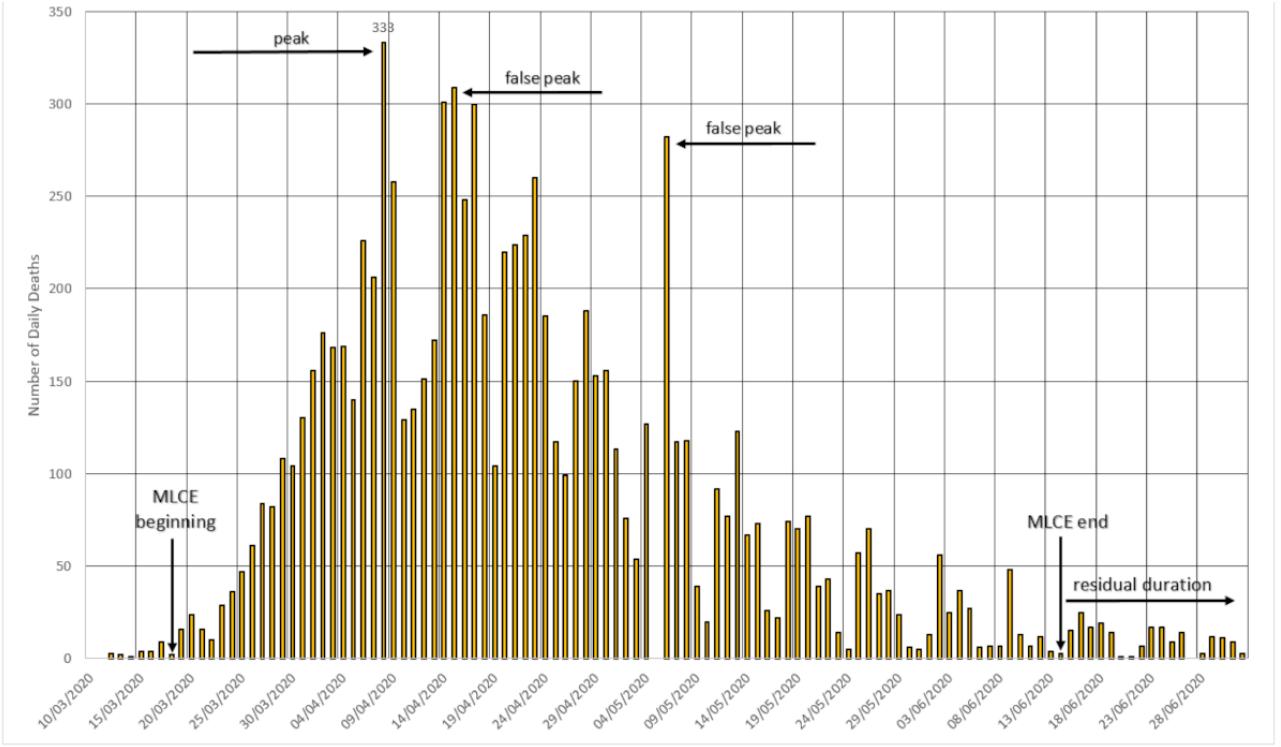
intermittently, but relatively small and oscillating, decreasing to certain levels of daily deaths, where it then becomes apparently chronic and presents relatively low values, but remains greater than zero.

In order to show numerical cases of the application of the proposed model, data from three European countries with different cycles were analyzed: Germany, a country that was reported as exemplary in terms of application of nonpharmaceutical interventions (NPIs); Italy, which stayed at the center of the initial crisis; and Sweden, which generally did not apply any strong NPIs. The data for this part of the study were obtained from the Worldometer's COVID-19 portal [15] as of July 9, 2020, and is presented, together with the calculations, in Multimedia Appendix 1.

Germany

Described from the beginning of the pandemic as a country that managed the crisis in an exemplary way, testing significant portions of its population and controlling and lifting restrictions on public movement based on well-known numbers and percentages of cases. Figure 1 shows the evolution of deaths in Germany. This framework points to the existence of the so-called false peaks. These are local maximums that were recorded during the cycle of rising or falling in the trend of deaths, but they are not inflection points. In order for a point to be considered as a (real) peak, it is necessary to register a tendency of decline in the number of deaths. This fall will not be linear, but there is an obvious, numerical, and visual trend that indicates such a pattern.

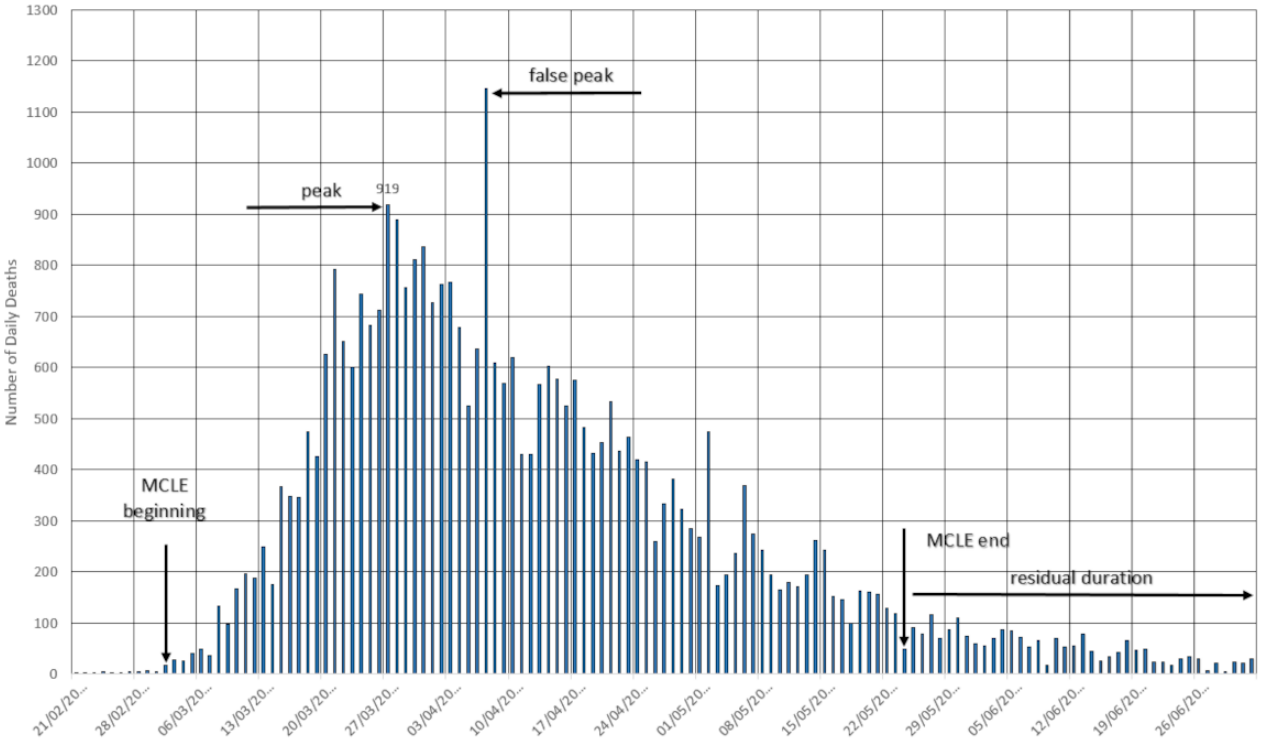
Figure 1. The cycle in Germany. MLCE: most lethal cycle of the epidemic. Source: Worldometer [15].



Italy

A country that was at the European epicenter of the crisis, Italy experienced an evolution in the number of deaths (Figure 2), which indicates the overcoming of the MLCE.

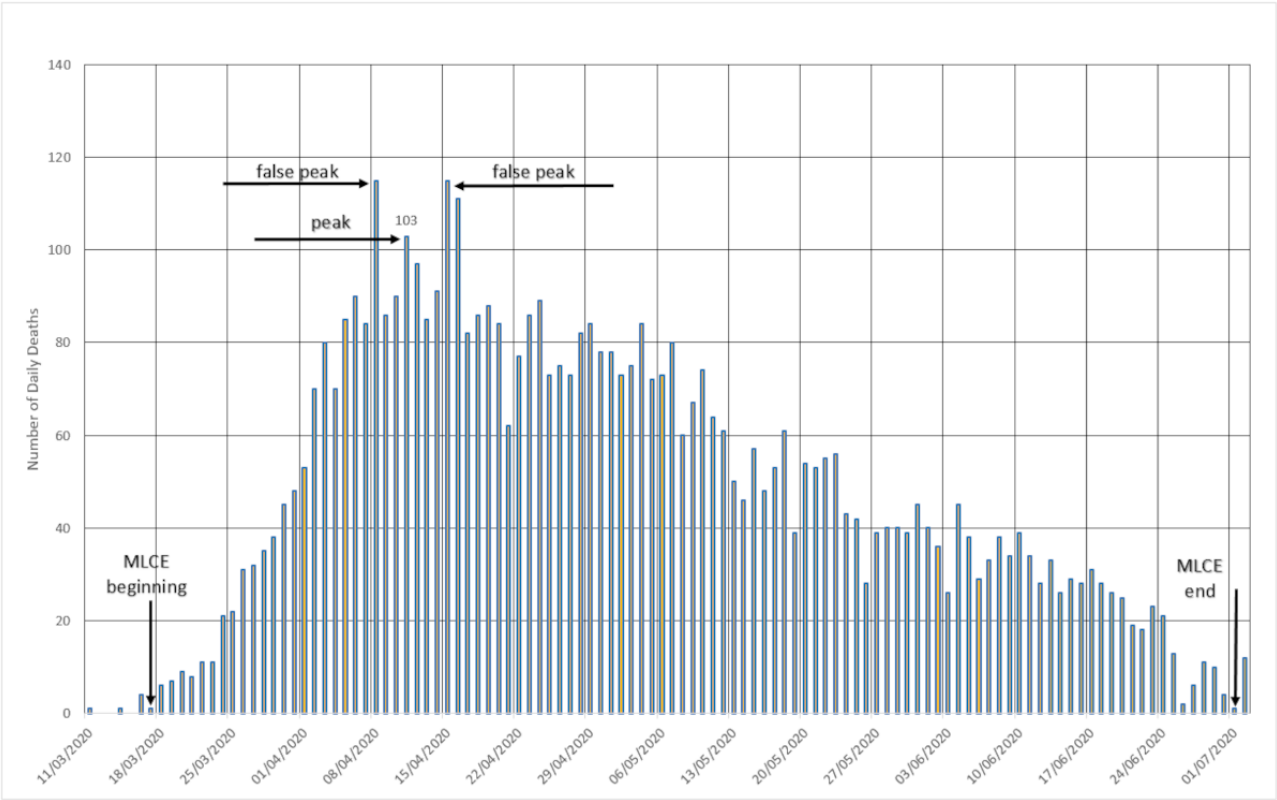
Figure 2. The cycle in Italy. MLCE: most lethal cycle of the epidemic. Source: Worldometer [15].



Sweden

Sweden, an European country that has not adopted the practices of radical social isolation like its neighbors, has a cycle of aspect not unlike that of all other European countries. Figure 3 shows the values of deaths that have already been corrected for the dates on which they actually occurred and not the date of registration.

Figure 3. The cycle in Sweden. MLCE: most lethal cycle of the epidemic. Source: Worldometer [15].



Nondimensional Characteristics of Epidemic Cycles

In general, the epidemic cycles described here have some common geometric characteristics, the main one being a triangular aspect (Figure 4), where a smaller side is formed, which corresponds to an average daily increase in the number of deaths until a peak is reached. This peak may be easily identifiable or require extrapolation of a line because the values oscillate naturally and some spurious points (false peaks) may appear. The peak is followed by a period where the number of deaths occurring daily tends to decrease on average. This period, for the observed cases, is longer than the previous one. According to Kotz and Rene van Dorp [16], the triangular distribution is used when there is no exact idea of what the distribution is, although there is an idea of the minimum and maximum values for the variable. Therefore, this distribution was chosen given its particular nature and use in situations where the description of a given population is uncertain, as is in this case. This distribution is based on the minimum and maximum estimates. Hence, Table 1 gathers values of the so-called triangular cycles presented earlier.

The values listed in Table 1 indicate that the period of rise of the disease in countries of relatively small sizes or in big cities is about 21 days, ranging from 19 to 25 days before reaching the so-called peak. From then until the end of this critical period, about 60 days pass, ranging from 45 to 81 days. The ratio between the two periods oscillates between 2.1 and 3.3, with an average of 2.8. Table 2 shows the number of deaths in the periods described above.

Figure 4. The generic shape of COVID-19 lethal cycles.

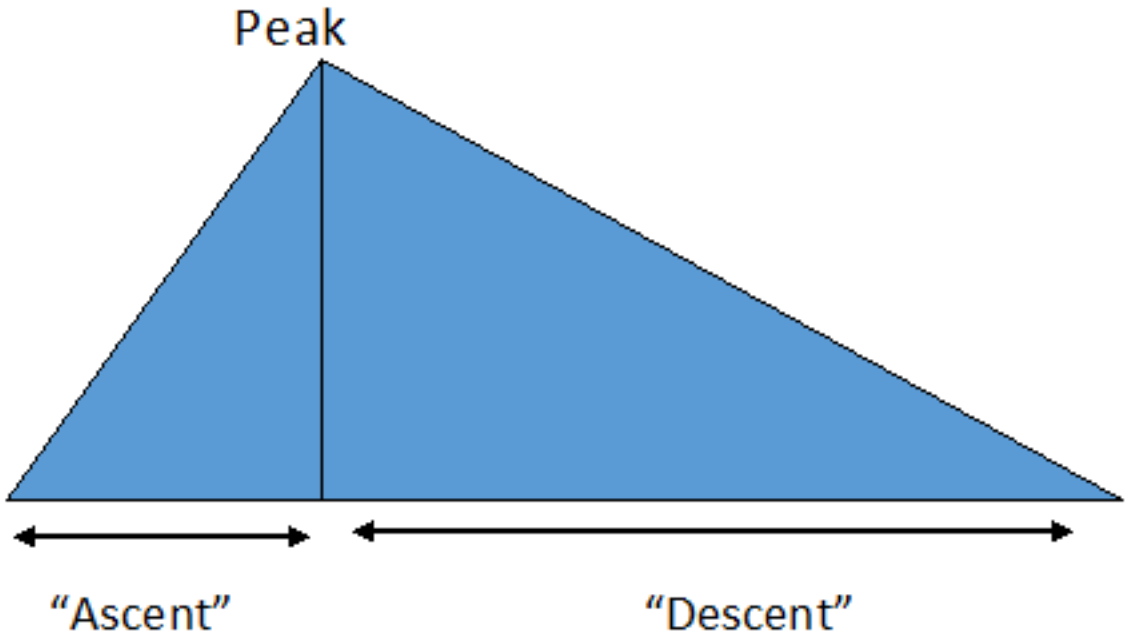


Table 1. Proportions between the time of ascent until the peak of deaths and descent to the end of the most severe cycle of COVID-19.

Country	Start	Peak	End	Days to the peak	Days to the end	Proportion between ascent and descent
Italy	March 7	March 27	May 24	20	57	2.9
Sweden	March 17	April 11	July 1	25	81	3.2
Germany	March 18	April 8	June 14	21	69	3.3

Table 2. Proportions between the number of deaths associated with the cycle of rising to the peak and of descending to the end of the most severe cycle of COVID-19.

Place	Start	Peak	End	Deaths to the peak	Deaths to the end	Proportion between ascent and descent
Italy	March 7	March 27	May 24	8937	24,082	2.7
Sweden	March 17	April 11	July 1	1255	4141	3.3
Germany	March 18	April 8	June 14	2323	6521	2.8

The values listed in Table 2 indicate that the number of deaths during the period of ascent of the disease in countries of relatively small sizes or cities is about 5791 (range 1255-10,293) before reaching the peak. From then until the end of this critical period, about 12,673 (range 4141-24,082) deaths occur. The

ratio of death figures ranges from 1.6 to 3.3, with an average of 2.4.

Therefore, it is possible to identify that once the scale effects are removed, what remains is a spectrum of proportions of the epidemic cycle. Then, when submitting the data to the moving average method with the initial value (MAMI), there is a minimization of the effect of seasonality in the registration of deaths, caused by weekends, holidays, and other local peculiarities. After dividing all the values previously transformed by the peak of the series (peak now determined by MAMI), the values start to be dimensionless and fall between 0 and 1. In this way, the epidemic cycles can be compared with each other, since what remains are the proportions between the ascent, the peak, and the descent of the cycle. The time period does not change. One clear limitation of this method is the necessity of identifying the real peak. Then, a hypothesis arises that different locations may, under different behavioral rules, present the same behavior.

Algorithm for Cycle Predictions

After identifying the triangular pattern and through successful application in several cases, a prediction algorithm was developed, described by the following steps:

1. MAMI is calculated for the daily figures on the number of deaths.
2. The set of values is normalized and MAMI is also applied on that.
3. A continuous curve is generated on a graph with the x axis as the number of consecutive days of the epidemic cycle and the y axis as the dimensionless range from 0 to 1 (some points, the false peaks, can go beyond this).
4. Among countries or localities, we seek those that have already ended their critical epidemic cycle (MLCE) and that are visually similar to the curve obtained in step 3, although obviously on a different scale, becoming the locality of reference.
5. MAMI is applied to the locality of reference.
6. Data of the locality of reference are normalized.
7. Repeat step 3 for the data of the locality of reference.
8. Considering that the cycle of the locality of reference is finished, it will be positioned previously on the graph, in relation to the place where it is desired to estimate the probable end date of the critical cycle. One should then numerically superimpose the peak of the case under study with the reference.
9. Once the superposition is made, always moving the reference case, an extrapolation can be made using the reference case as a guide to the value to be determined. As the scale of the case studied has not been changed, it is enough to consult what day it would be in the future to know the probable date.
10. If there is no similar case, you can eliminate the last days, as discussed above, and extrapolate directly from the values obtained in the public databases.

Effective Reproduction Number

After identifying the similarities between cycles, the next step is to calculate the R_t, which is done on the experimental behavior of the curve. First, however, it is necessary to understand the effect of MAMI on the reproduction number.

MAMI Effect on Reproduction Numbers

The impact of MAMI applied to registered numbers can be better understood by analyzing Figure 5, where MAMI bears the greatest effect at the very beginning of the epidemic cycle; however, after a brief period, the average and actual data tend to yield to the same value as the cycles progress. It will be shown along this paper that the reproduction number varies most in the early stages, and the use of MAMI is plainly justified to avoid numbers that are registered in batches and not into a smooth daily fashion. Daily figures for total cases collected from the Johns Hopkins University's website [17] on July 22, 2020, together with the calculations, are presented in Multimedia Appendix 2. The analysis of the R_t for the three European countries are represented in Figures 6-8.

Figure 5. MAMI (moving average method–initial value) effect on reproduction numbers (R_t) expressed for two different countries, South Korea (SK) and Italy. South Korea: the blue line is R_t obtained from MAMI applied to registered data; the red line is R_t determined for registered data. Italy: the yellow line is R_t for registered data; the green line is for MAMI applied to registered data. Source: Johns Hopkins University [17].

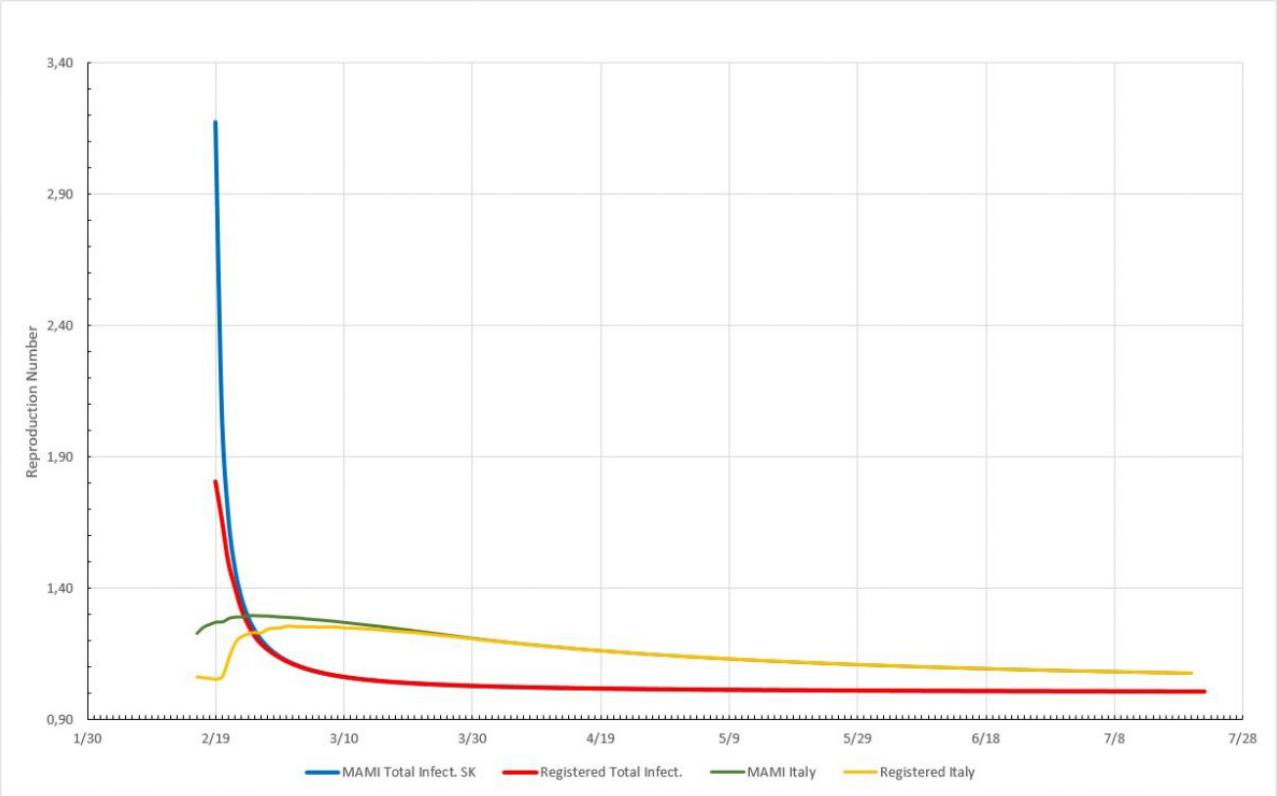


Figure 6. Number of COVID-19 cases reported for Germany. The black line represents the daily reported numbers, the blue bars their MAMI (moving average method–initial value), and the red line the total cases to date, using the right-hand axis as reference. Source: Johns Hopkins University [17].

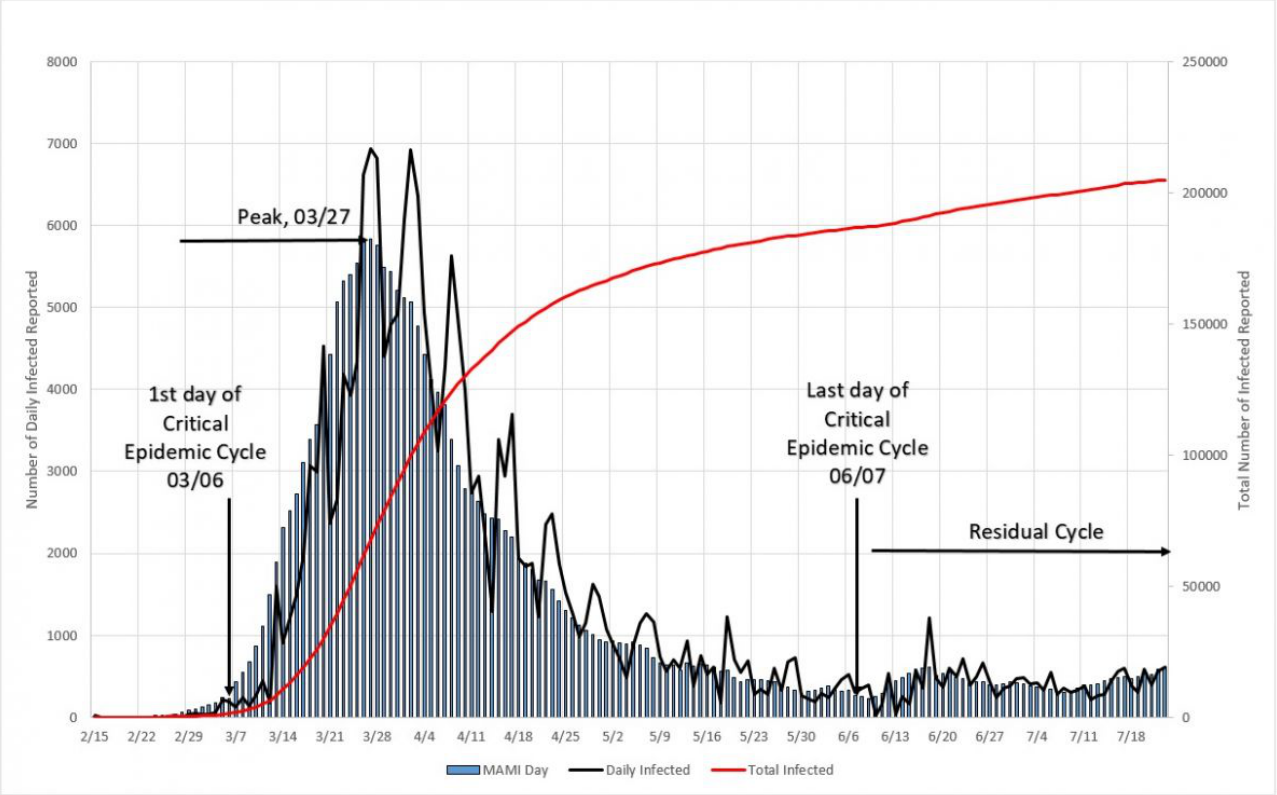


Figure 7. Number of COVID-19 cases reported for Italy. The black line represents the daily reported numbers, the blue bars their MAMI (moving average method–initial value), and the red line the total cases to date, using the right-hand axis as reference. Source: Johns Hopkins University [17].

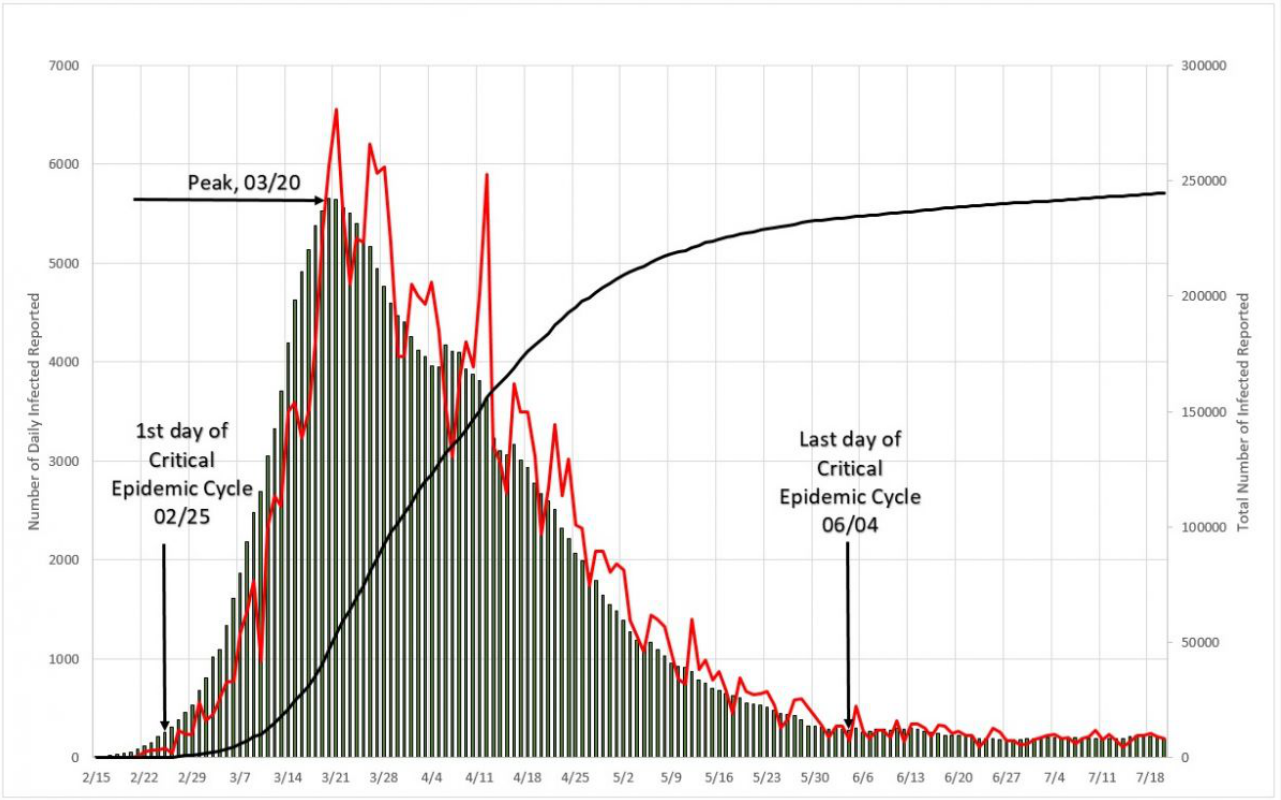
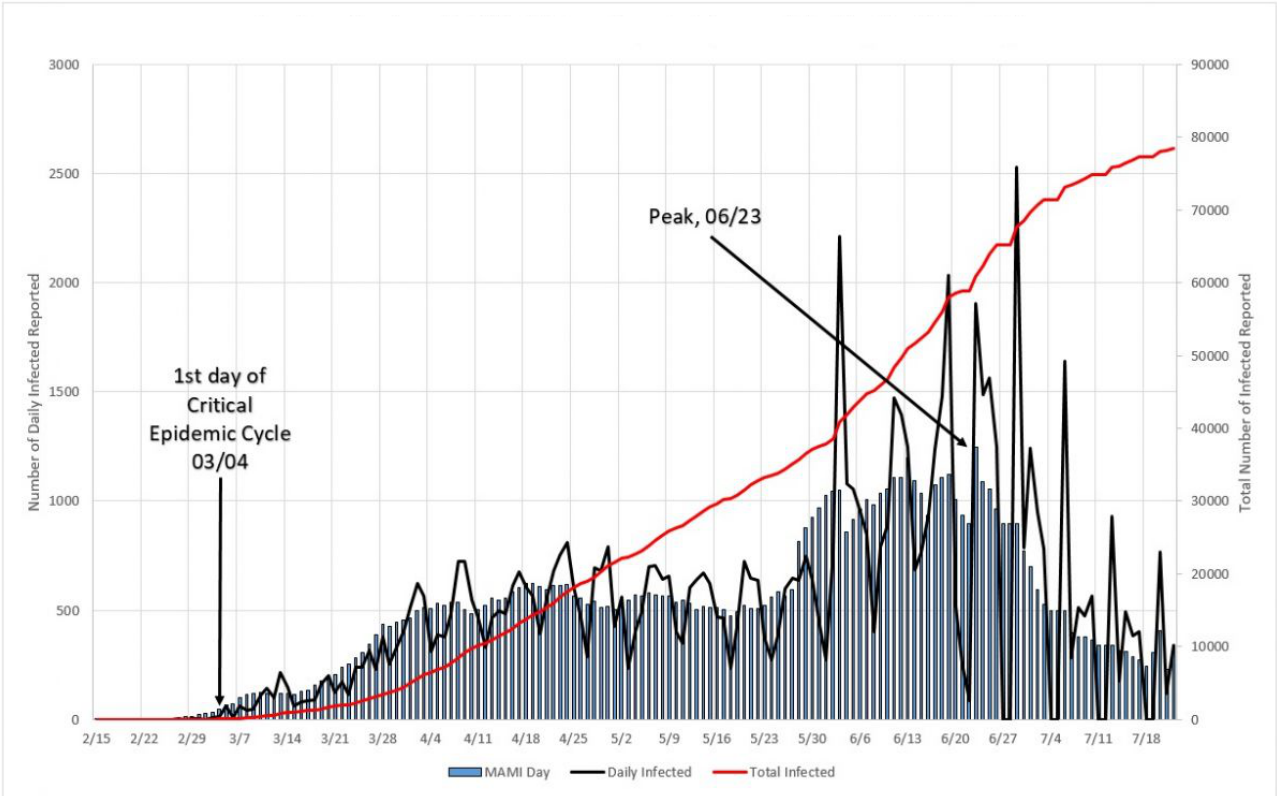


Figure 8. Number of COVID-19 cases reported for Sweden. The black line represents the daily reported numbers, the blue bars their MAMI (moving average method–initial value), and the red line the total cases to date, using the right-hand axis as reference.



Deriving the Effective Reproduction Number

With the effect of moving averages measured, it is possible to proceed to an experimental method for calculating the daily number of infected and then an effective, time-varying reproduction number, calculating its value by means of experimental data outlined below.

The total number of infected daily (I_d), during a period of time t, can be described as a function of the daily increase rate factor ($1+b$) multiplied by a scale factor, as shown in equation 1:

$$I_d = a\,(1+b)^t \quad \textbf{(1)}$$

In equation 1, a is the scale factor and b is the absolute daily increase rate, or instantaneous rate, and is defined as:

where $I_{d,n+1}$ is the current day and $I_{d,n}$ is the previous day.

Equation 1 can be written as:

$$I_d = C^t \quad \textbf{(3)}$$

where C is the time-dependent effective reproduction number, $R_e(t)$, or R_t for short, which is obtained from experimental data. For the reproduction number determination, it is necessary to determine the scale factor a. Therefore, a takes the following form:

Finally, from equations 3 and 4:

In order to map the interpretation proposed from equations 1 to 5 to the classical mathematical interpretation for the reproduction number (R_0), an equivalence transformation will be described as follows. From the classical definition of R_0, let:

where β is infection-producing contacts per unit time (instantaneous rate), with a mean infectious period of τ. Equation 6 can be transformed into:

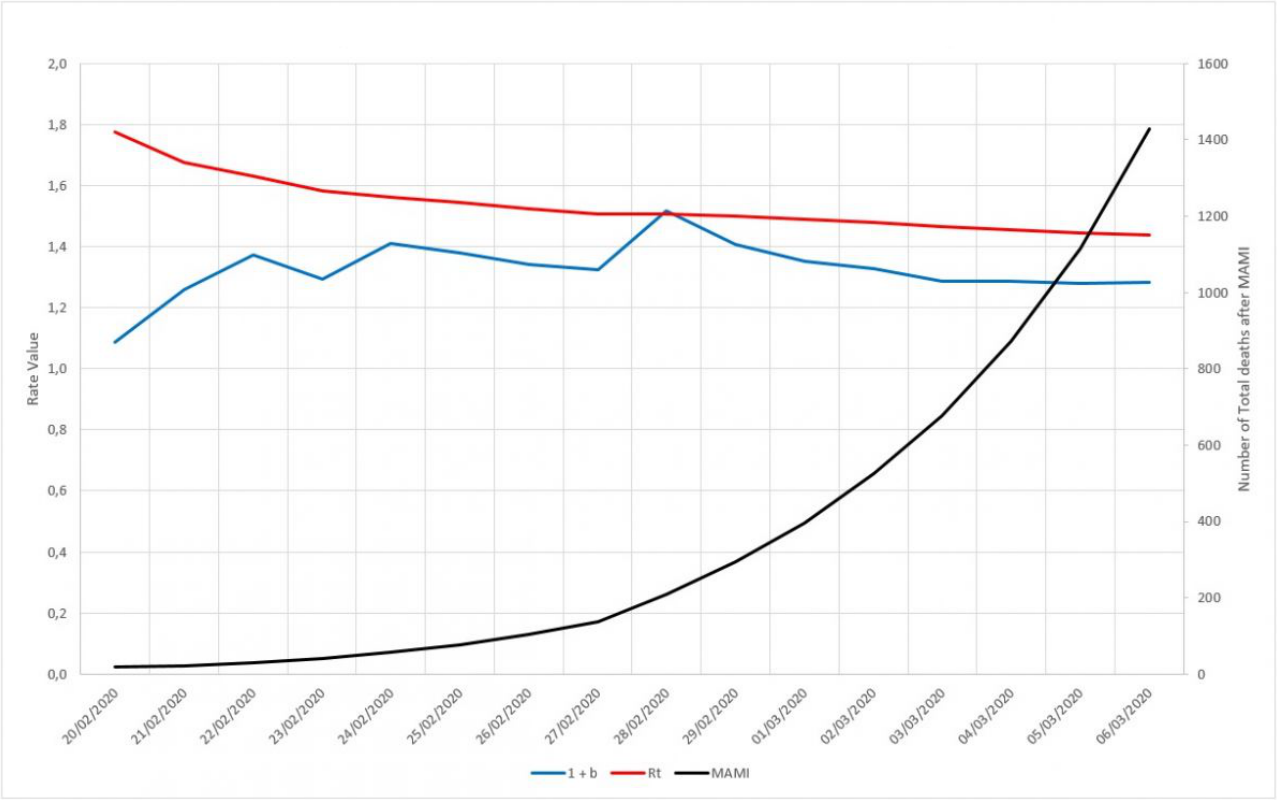
$$R_0 = e^{k\tau} \quad \textbf{(7)}$$

From equations 5 and 7:

In equation 8, all dimensional units are compatible, therefore our transformations to obtain R_t in equation 5 are valid. Equation 5 was obtained from experimental data, and it is at the core of the model proposed here. From this point onward, R_t must be interpreted as $R_e(t)$ as explained before, in the interpretation of equation 3.

During the data analysis, we noted that the daily increase rate factor ($1+b$) is not enough to describe the number of contaminated cases registered in a given day, because it simply informs the absolute increase ratio that occurred from one day to the next. The reproduction number coefficient needs more numerical information in order to be able to express correctly the magnitude of daily numbers. It needs the scale factor a to bring more information on the phenomenon. As an example of this finding, Figure 9 shows that while the ($1+b$) factor varies rapidly, R_t drops steadily, changing slowly as the exponential time grows. The same behavior is displayed by the total daily registered number of deaths, which keeps growing smoothly. This is the numerical evidence that the factor ($1+b$) alone cannot describe the total number of deaths.

Figure 9. Behaviors of (1+b) and effective reproduction number (Rt) factors for the first 20 days in the epidemic cycle of Germany. MAMI: moving average method–initial value.



Subnotification Effect on the Reproduction Number

When it comes to analyzing the number of cases of infection in the COVID-19 epidemic, an issue that always arises is underreporting or subnotification and its importance in predicting the behavior of the epidemic cycle. Thus, the third part of the framework is dedicated to the study of subnotification and its effects on prediction. Subnotification is understood as the fact that counts of infected persons are only estimated by public health authorities. Given that many people exposed to the virus do not display any sign of infection or the symptoms are very mild, therefore going unnoticed and unregistered by local bureaus of health statistics, the development of evaluation tools of the impact of these nonnotified cases is necessary. If it is assumed that subnotification is a constant factor (eg, 10 times the registered number of cases) during the whole epidemic cycle, it does not change the absolute daily increase rate *b* or the (1+*b*) factor. However, it does affect the scale factor *a*, therefore changing R_t.

Subnotification Impact Estimation Method

The impact of subnotification on R_t may be estimated by initially assuming that the actual registered figures for daily infected persons are no longer their actual values, but "real" ones multiplied by a factor—the subnotification factor. After that, the scale factor *a* is calculated. The term (1+*b*) remains constant, once the ratio (equation 3) remains constant. Then *a* and (1+*b*) are applied to equation 5, thus recalculating R_t, now reflecting the effect of the imposed subnotification factor. This new R_t value would have been the correct one, in case all subnotified cases were suddenly registered. The percentage difference between this new, recalculated R_t and the actual one provides an estimate for the impact of subnotification on the reproduction number for a given population. Therefore, multiplying the values for registered cases by a factor of 10 will not cause a tenfold increase in R_t. The true impact must be therefore calculated as described. It is also observed that subnotification mostly affects the very beginning of the critical cycle. After a certain amount of time, errors drop to insignificant values, below 5%.

Total Number of Infected, Daily Infection Rate, Lag Time, and Incubation Period

The fourth component of the framework is the application of the logistic model to estimate three parameters: the total count of infected individuals; the daily infection rate; and the lag, which defines when the cycle actually started. An innovative model, based on the concept of inventory formation, is used to determine a fourth parameter—the most likely incubation period for the virus.

Considered by many authors as a good fit for modeling epidemic episodes [18-20], the logistic model describes three typical phases for this type of episode: the slow start, the steady growth, and finally the asymptotic behavior of the end. There are several ways to implement this function, and this work will use the so-called Richard growth model to describe the accumulated number of infection cases. The generalized logistic function has the following form:

By selecting the highest r^2 among several variations of equation 9, through curve-fitting, a particular form for equation 9 is:

where $N(t)$ is the number of infected persons at a given period of time t, a is the final count for the total infected, b is the daily infection rate, c is the lag phase, and d is a positive real number. It can be shown that:

The constants a, b, c, and d will be used to estimate x_1, the maximum number of infected people in a given location; x_2 is the daily infection rate, or the average absolute daily increase in the number of infected, which can be used to determine the reproduction number (and to estimate the incubation period). Finally, x_3 is used to estimate the lag time, or the actual moment when the first case occurred.

Incubation Period Estimation

Although there is a series of studies on the incubation period for SARS-CoV-2, in order to maintain consistency within the framework, we sought to develop a model that could also estimate what would be the best incubation period estimation method to consider when modeling epidemic cycles. For that, we defined a model of inventory of infected people similar to the one used in productive systems, as shown in equation 12:

$$I_t = I_{t-1} + D_t - D_{t-n} \quad \textbf{(12)}$$

where I_t is the inventory of people infected in day t, or the total of infected in day t; I_{t-1} is the inventory of people infected in the previous day; D_t is the number of people detected with the disease in day t; and D_{t-n} is the number of people detected with the disease n days before t.

Equation 12 should be interpreted as follows: the number of people who are infectious on a given day is equal to the number of people who were infectious the day before, plus the number of infected detected on the same day, and minus the number of people who have left the N-day incubation period. This reasoning therefore assumes that as soon as a person finds out he or she is infected, that is, when this person leaves the incubation period, enters perfect isolation and stops infecting. Although this assumption is not completely realistic—since it depends not only on individual responsibility, but also on the implementation of efficient isolation measures—at the same time it must also be considered that not every infected person effectively infects others, given that isolation is not the only way to avoid viral contamination. Thus, we consider this assumption to be reasonable enough to be applied statistically.

Other basic assumptions are that of all people susceptible (not vaccinated, sufficiently exposed to the pathogen, etc), not all will expose or develop the disease in a form severe enough to be noticed. Accordingly, the recorded number of daily cases does not reflect the total number of infected, but those who seek medical attention and therefore were diagnosed as contaminated. Hence, this is the number of infected in a given day, or the "inventory" of people that can infect other people in a given day. With the formulation defined in equation 12 and the assumptions described previously, we carried out the analysis and simulations for the three countries.

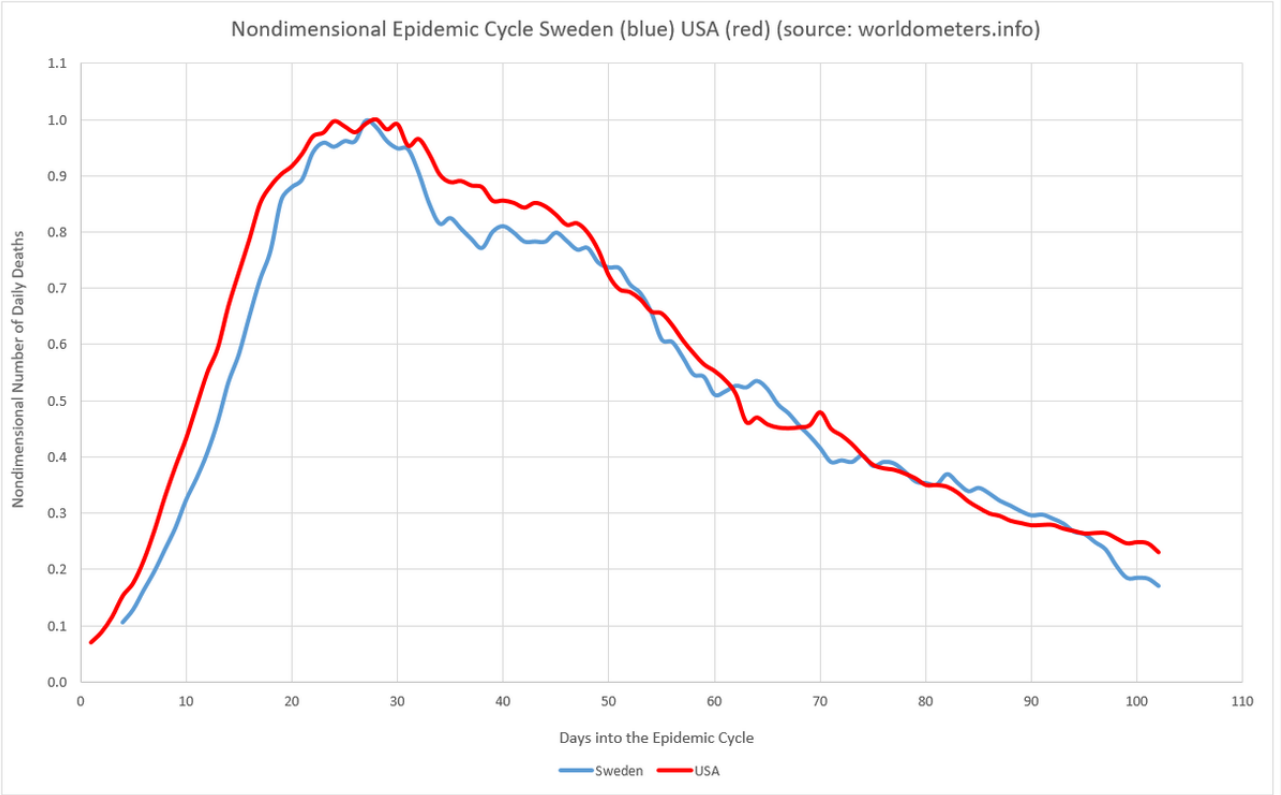
Results

General Findings

The epidemic cycles observed were subjected to the numerical methods present in the framework and described in the previous section. The first data transformation was the application of the MAMI value. The second transformation was normalization, where all the values were divided by cycle peak value, causing most of the values to fit between 0 and 1, except for the false peaks. These two consecutive transformations allowed for a comparison of behaviors among cycles and proved that several epidemic cycles, within the pandemic, have similarities. With these first steps, it is possible to estimate the duration and general behavior of a local episode, even though this, in absolute terms, does not present the same number of deaths or duration as a similar cycle. What remains approximately constant are the proportions of similar cycles. This technique has been applied with great success in the performance prediction of professional athletes and teams [21].

By the time the analyses were done, the three countries considered in this paper presented more advanced cycles, so no predictions were made for them; instead, their cycles were used to perform analysis on other countries, regions, and cities. For instance, Figure 10 presents the similarity of the United States' and Sweden's cycles. A complete set of predictions for Brazil, the state of Rio de Janeiro, and the city of Rio de Janeiro, as well as a measurement of the performance of the model, are presented in Multimedia Appendix 3. In addition, as seen in De Carvalho and De Carvalho [12], it is possible to find many other comparisons and predictions between cities, regions, and countries using this method.

Figure 10. Comparison of epidemic cycles: Sweden and the United States. Source: Worldometer [15].



The analyses of the other variables considered in the framework for Germany, Italy, and Sweden are presented in the next sections. The data for this part of the study were also collected from the Johns Hopkins University's website [17] on the declared dates.

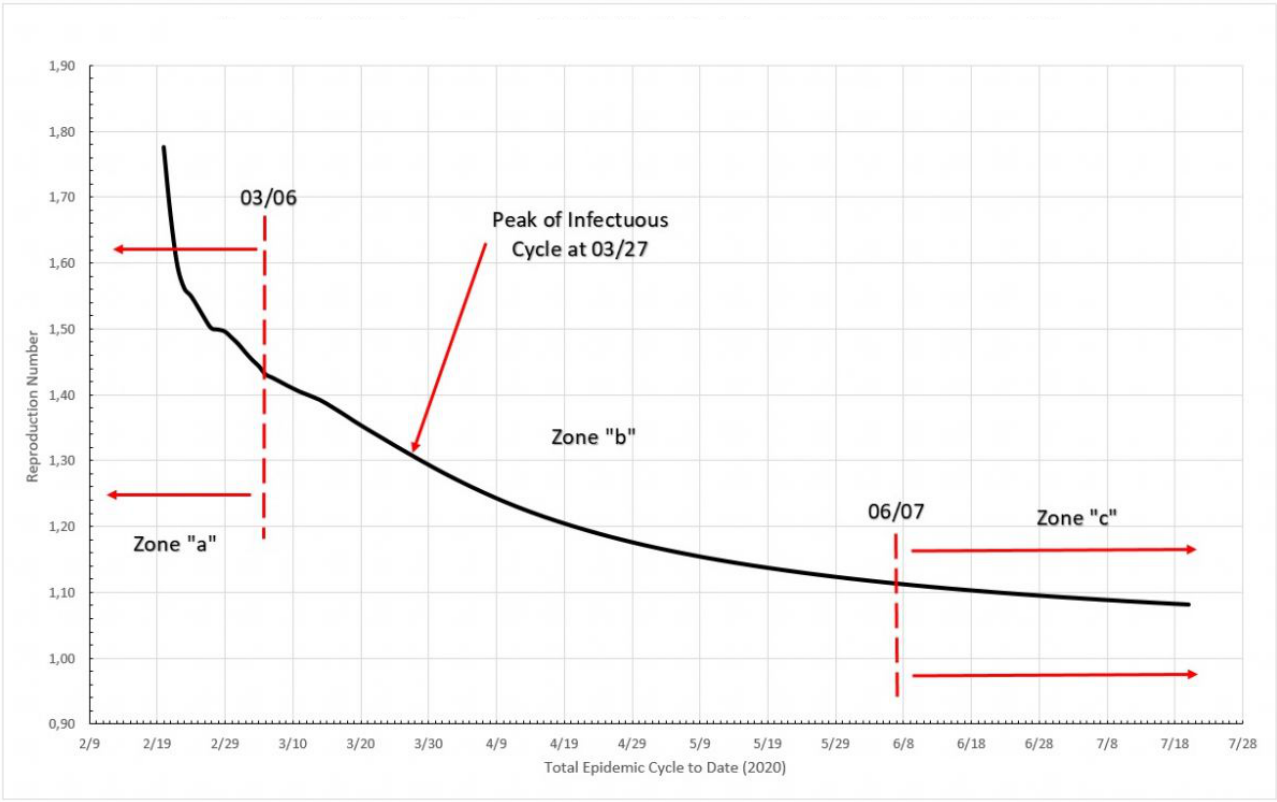
The expressions developed in equations 1 to 5 do not explicitly take into account the incubation period, with the instantaneous rate of change, or daily increase in number of registered infected individuals, calculated as defined in equation 5. For the sake of thoroughness, three simulations were performed, for an incubation period of 5, 10, and 15 days. This was achieved by redefining the expression (1+*b*) for a new set of parameters, basically dividing the total number of reported cases for a given day by the values registered in 5, 10, and 15 days before. In that way, the term (1+*b*) would now reflect the incubation period over R_t. All simulations yielded zero (0%) change, to the fourth significant figure. Therefore, it is assumed that the described method is inherently insensitive to incubation period variations or influence, reinforcing its simplicity and robustness. The data and calculations are in Multimedia Appendix 4.

Germany

Reproduction Numbers

In Figure 11, three distinct zones are formed. Zone "a" is in the very beginning of the cycle, and the reproduction number varies from 1.10 to 1.48 from one day to the next; this is probably only the reflection of large initial variation in numbers but only if we limit this zone to no more than 5% of the MAMI peak value. It is easy to notice that the figures bear small influence on the overall disease behavior. Zone "b" describes the transmission during the critical disease cycle (from March 6 to June 7), where a rapid increase in daily cases stops only around the peak than drops steadily toward the end. This is the most lethal period of the epidemic cycle, and it is considered over once a 5% peak level is reached again. The remaining time, zone "c," is the residual cycle that appears in all countries and places facing the COVID-19 crisis. In absolute values, the reproduction number for the critical period starts with a value of 1.30 and drops continuously toward 1.00, although never quite reaching it (at the time this paper was written).

Figure 11. Total epidemic cycle in Germany, using the daily number of infected people. Source: Johns Hopkins University [17].



Subnotification

An arbitrary threshold line representing a 5% error was drawn in Figure 12. This limit shows that after the 50th day into the German critical cycle (the one between 5% of the peak value, before and after it), regardless of the amount of subnotification, the error of the calculated reproduction number is no greater than 5%, as presented in Table 3. At the other extreme, a 3x subnotification essentially does not induce errors greater than 5% on the reproduction number, at any time during the critical cycle. A maximum error of 16.84% is estimated for the worst case scenario simulated here, a 40x subnotification, and the first day into the cycle. In overall, subnotification appears to have no significant impact in Germany's official infected numbers. Subnotification also seems to have more impact in the very beginning of a given cycle but becomes irrelevant toward the end.

Figure 12. Subnotification effect on reproduction number in Germany during the critical epidemic cycle. Source: Johns Hopkins University [17].

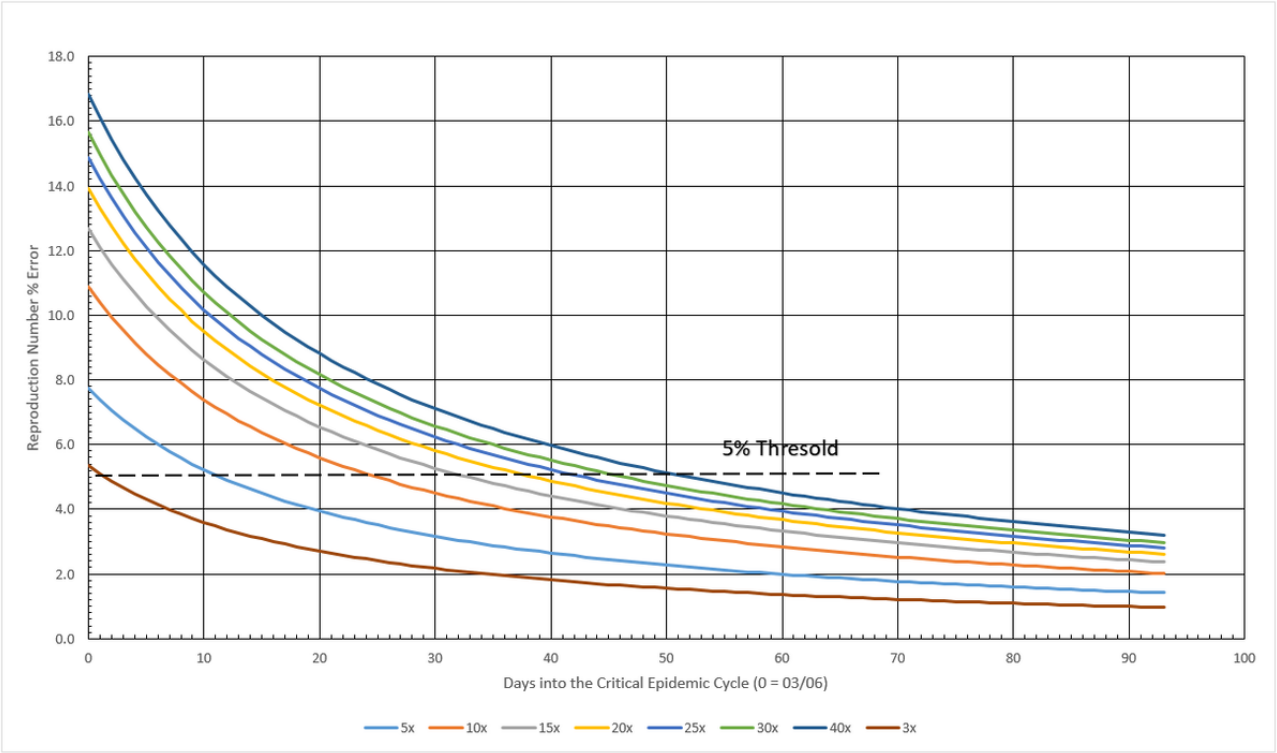


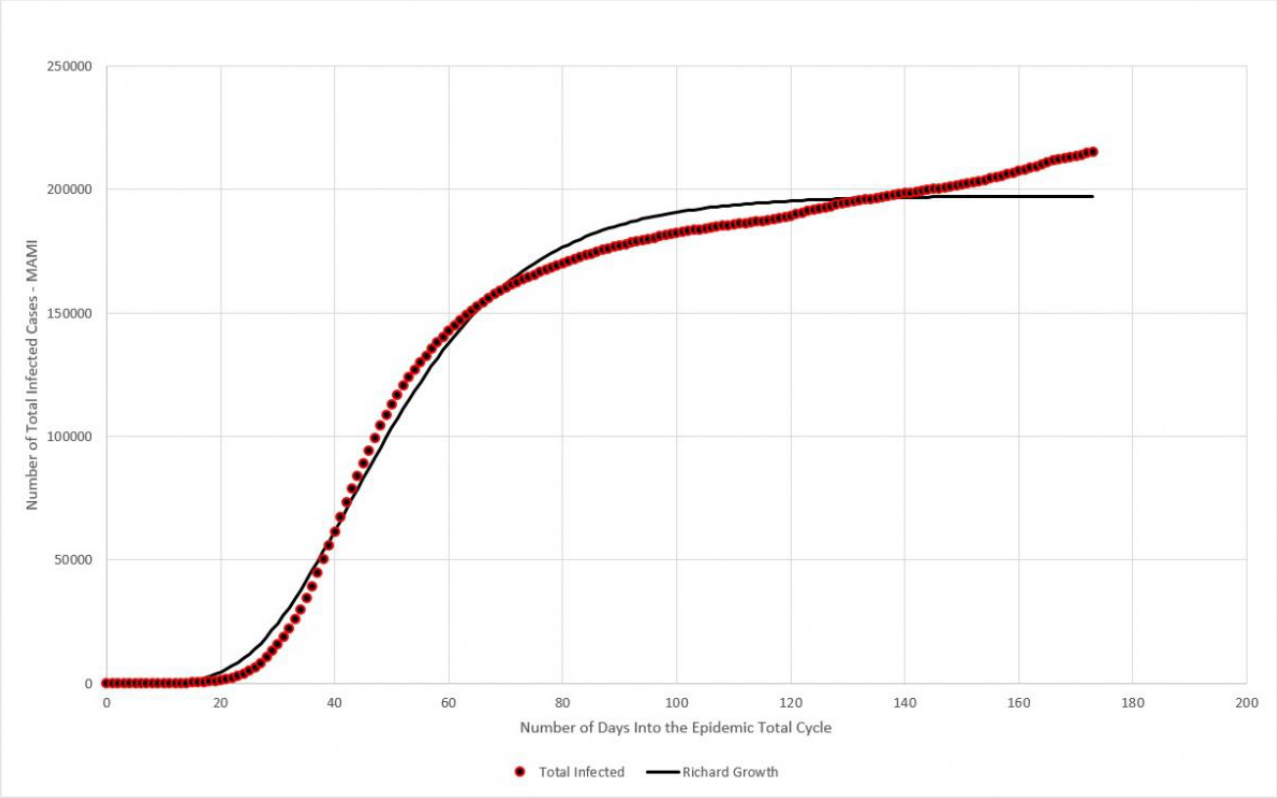
Table 3. Errors associated with ignoring the existence of subnotification in the epidemic cycle.

Subnotification	Max error (%)	Min error (%)	Days until ≤5%	Error (%) at peak day
3x	5.34	0.97	2	2.64
5x	7.73	1.41	12	3.85
10x	10.87	2.02	25	5.46
15x	12.66	2.37	33	6.39
20x	13.91	2.62	39	7.05
25x	14.87	2.81	43	7.55
30x	15.64	2.97	47	7.96
40x	16.84	3.21	52	8.60

Total Number of Infected

Data collected for Germany from February 15 to July 20 were plotted in Figure 13. The blue dots represent the daily registered infected cases submitted to MAMI, and the red continuous line represents the Richard growth model curve, drawn using parameters determined by the MAMI data.

Figure 13. Total number of infected (moving average method–initial value [MAMI]) compared to the Richard growth model prediction for Germany. Source: Johns Hopkins University [17].



As discussed previously, the German critical epidemic cycle started on March 6. Using curve-fitting data from Table 4, Table 5 shows that the first case must be recorded 89 days before that, with X_3 indicating that the first case of the total epidemic cycle occurred around December 8, 2019.

Table 4. Curve-fitting data.

Parameter	Value
a	197,372.97
b	–5.2260
c	0.0587
d	4.4208×10^{-4}

Table 5. Epidemic parameters determined using curve-fitting data from Table 4.

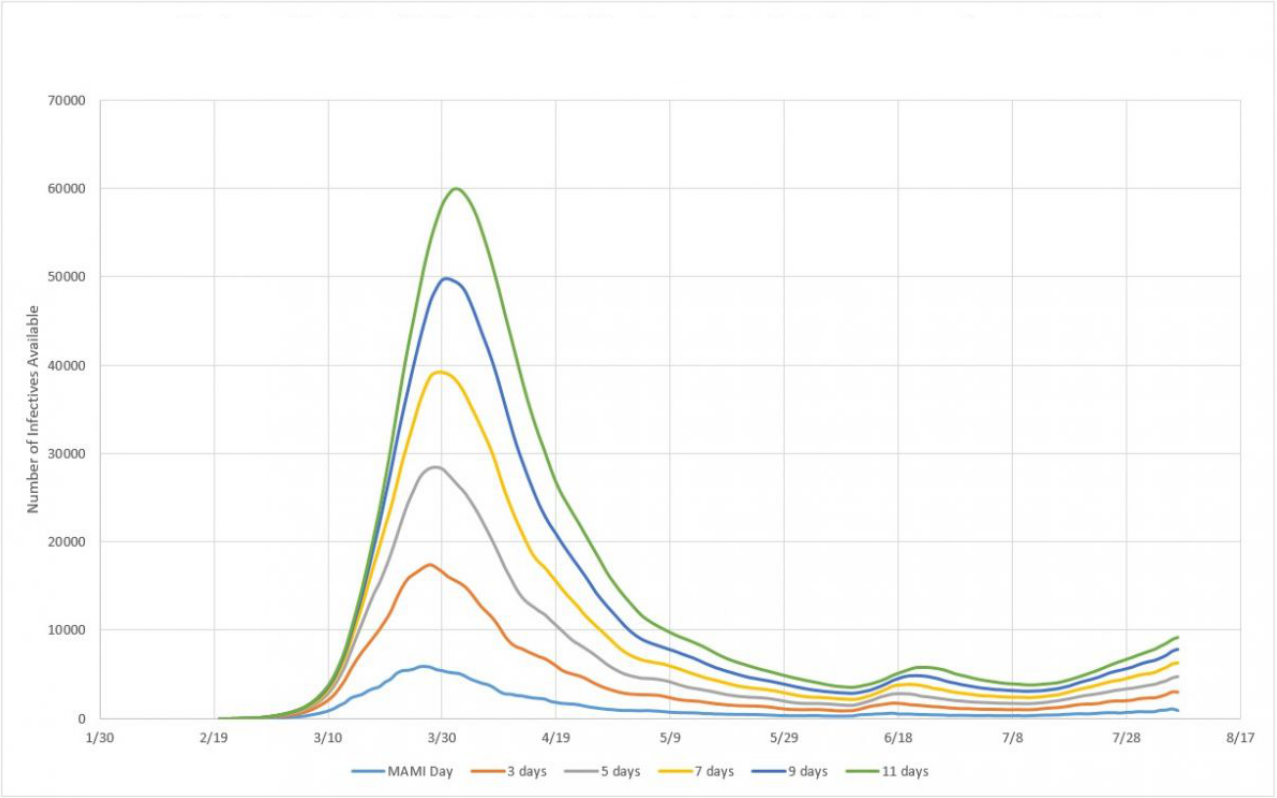
Epidemic parameter	Value
X_1	197,373
X_2	5.87[a]
X_3	89
r^2	0.9958

[a]Percent.

Impact of Incubation Period

In this section, we approach the model of formation of an infected persons inventory for the three countries considered. Simulations were made for incubation cycles of 3, 5, 7, 9, and 11 days. Inventories were calculated according to equation 12 and plotted together with the MAMI of detected cases. Figure 14 presents the subnotification study for Germany.

Figure 14. Infected person inventories for 3, 5, 7, 9, and 11 days of incubation, compared to MAMI (moving average method–initial value) for Germany. Source: Johns Hopkins University [17].

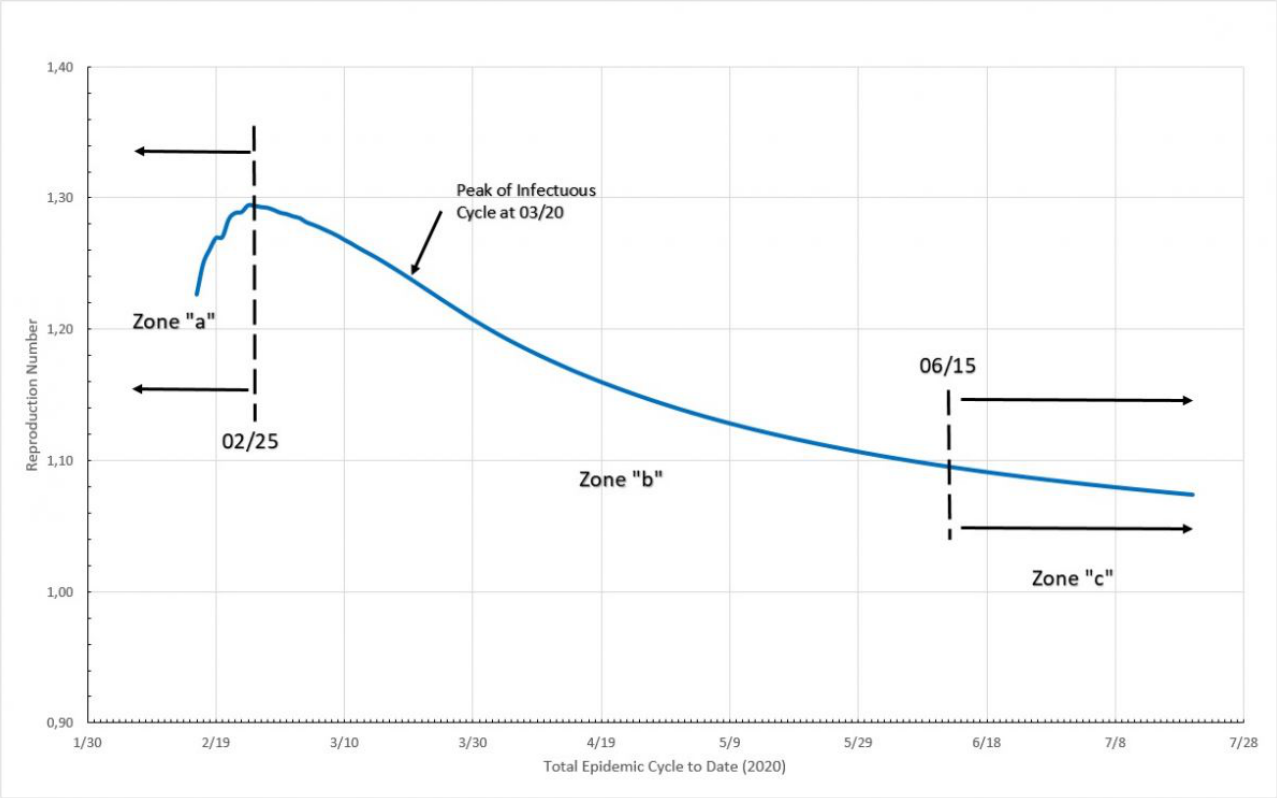


Italy

Reproduction Numbers

It can be seem in Figure 15 that three distinct zones are formed. Zone “a” is in the beginning of the cycle, and the reproduction number varies from 1.78 to 1.44 from one day to the next; once again this is probably simply the reflection of large initial variation in number, but this zone is limited to no more than 5% of the MAMI peak value. It is easy to notice that the figures bear small influence in the overall disease behavior. Zone “b” describes the transmission during the critical disease cycle (from February 25 to June 15). This is the most lethal period of the epidemic cycle, and it is considered over once a 5% peak level is reached again. The remaining time, zone “c,” is the residual cycle. In absolute values, the reproduction number for the critical period starts with a value of 1.44 and drops continuously toward 1.12.

Figure 15. Total epidemic cycle in Italy, using the daily number of infected people. Source: Johns Hopkins University [17].



Subnotification

Subnotification in Italy is presented in Figure 16. The 5% limit tells that after the 44th day into the Italian critical cycle, regardless the amount of subnotification, the error of the calculated reproduction number is no greater than 5%, as shown in Table 6. At the other extreme, a 3x subnotification essentially induces no errors larger than 5% on the reproduction number, in any time during the critical cycle, and 5x barely disturbs it. A maximum error of 12.34% is estimated for the worst case scenario simulated here, a 40x subnotification, and the first day into the cycle. Overall, subnotification appears to have no significant impact on Italy's official infected numbers, as in the previous two cases. Subnotification also has more impact in the very beginning of a given cycle but becomes irrelevant toward the end of it.

Figure 16. Subnotification effect on reproduction number in Italy during the critical epidemic cycle. Source: Johns Hopkins University [17].

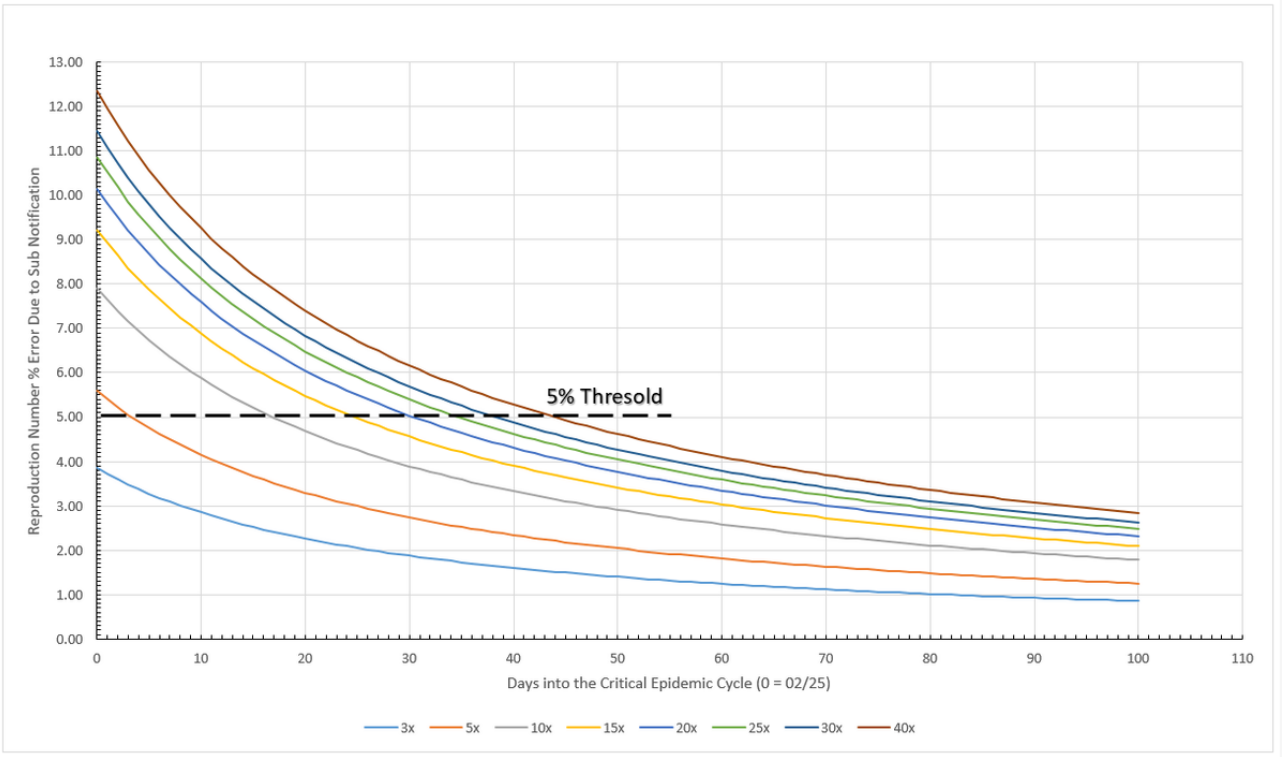


Table 6. Errors associated with ignoring the existence of subnotification in the epidemic cycle for Italy.

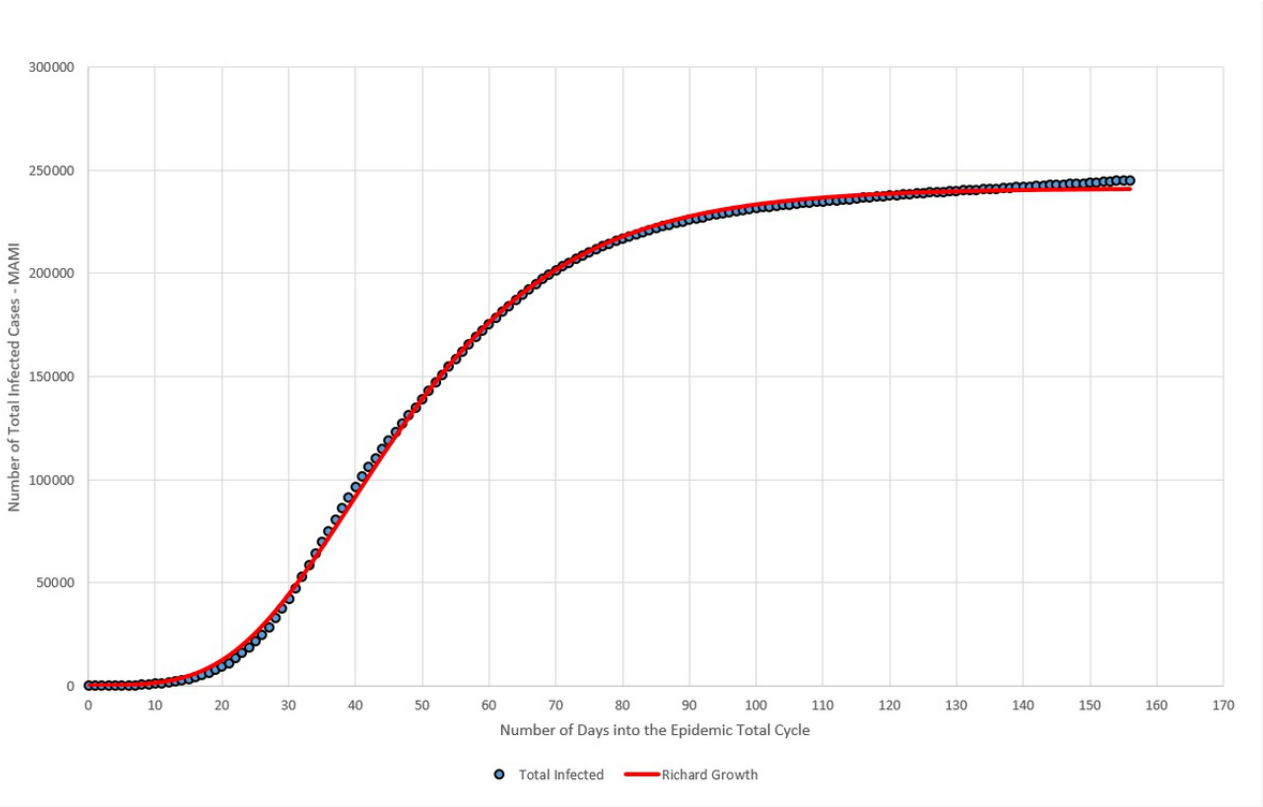
Subnotification	Max error (%)	Min error (%)	Days until ≤5%	Error (%) at peak day
3x	3.85	0.85	N/A[a]	2.09
5x	5.59	1.25	4	3.05
10x	7.89	1.78	17	4.33
15x	9.22	2.09	25	5.07
20x	10.15	2.31	31	5.60
25x	10.86	2.48	35	6.00
30x	11.44	2.62	39	6.33
40x	12.34	2.84	44	6.85

[a]N/A: not applicable.

Total Number of Infected

Data collected for Italy from February 15 to July 20 were plotted in Figure 17. The blue dots represent the daily registered infected cases submitted to MAMI, and the red continuous line represents the Richard growth model curve, drawn using parameters determined by the MAMI data.

Figure 17. Total number of infected (MAMI [moving average method–initial value]) compared to the Richard growth model prediction for Italy. Source: Johns Hopkins University [17].



The Italian critical epidemic cycle started on February 25. Using curve-fitting data from Table 7, Table 8 shows that the first case must be recorded 86 days before that, with X_3 indicating that the first case of the total epidemic cycle occurred around December 1, 2019.

Table 7. Curve-fitting data.

Parameter	Value
a	241,148.81
b	–4.8623
c	0.0562
d	8.4600×10^{-4}

Table 8. Epidemic parameters determined using curve-fitting data from Table 7.

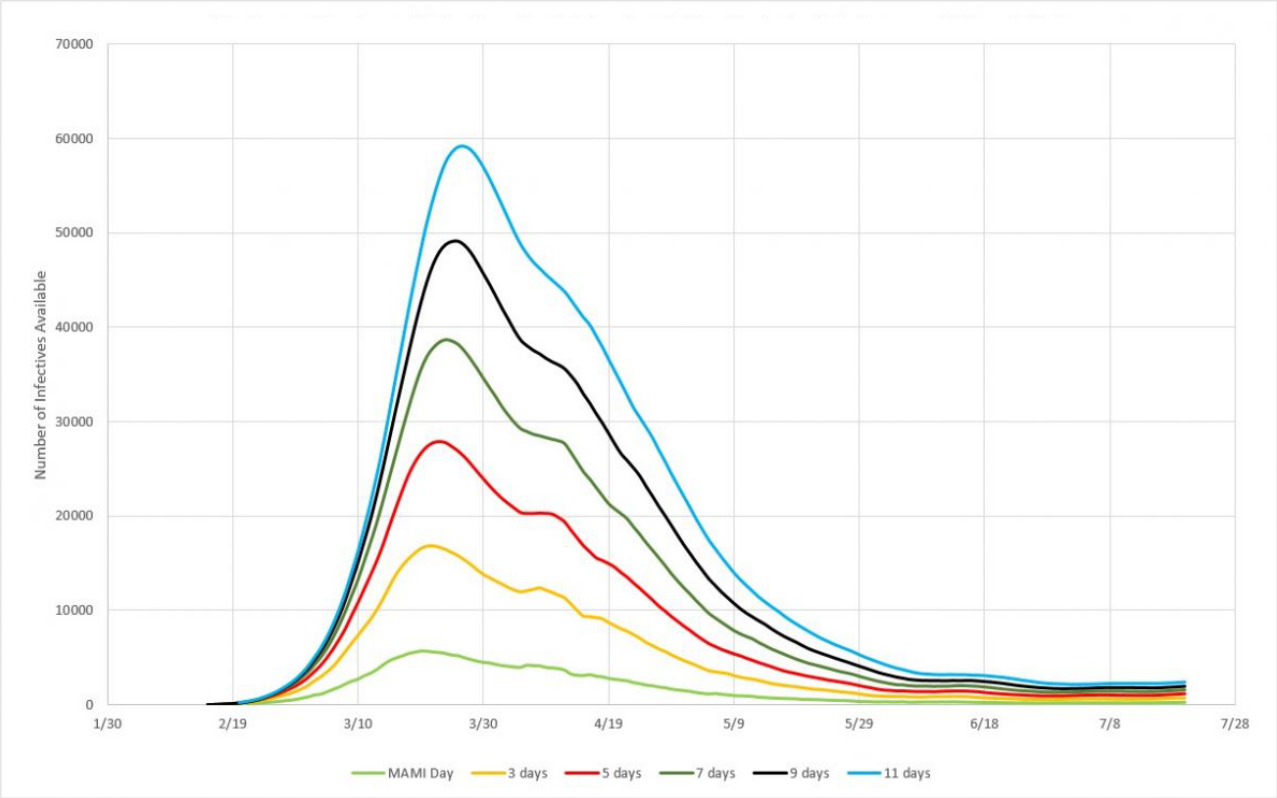
Epidemic parameter	Value
X_1	241,149
X_2	5.62[a]
X_3	86
r^2	0.9995

[a]Percent.

Impact of Incubation Period

Using the same reasoning applied to Germany, Figure 18 presents the inventories of infected persons for Italy.

Figure 18. Infected person inventories for 3, 5, 7, 9, and 11 days of incubation, compared to MAMI (moving average method–initial value) for Italy. Source: Johns Hopkins University [17].

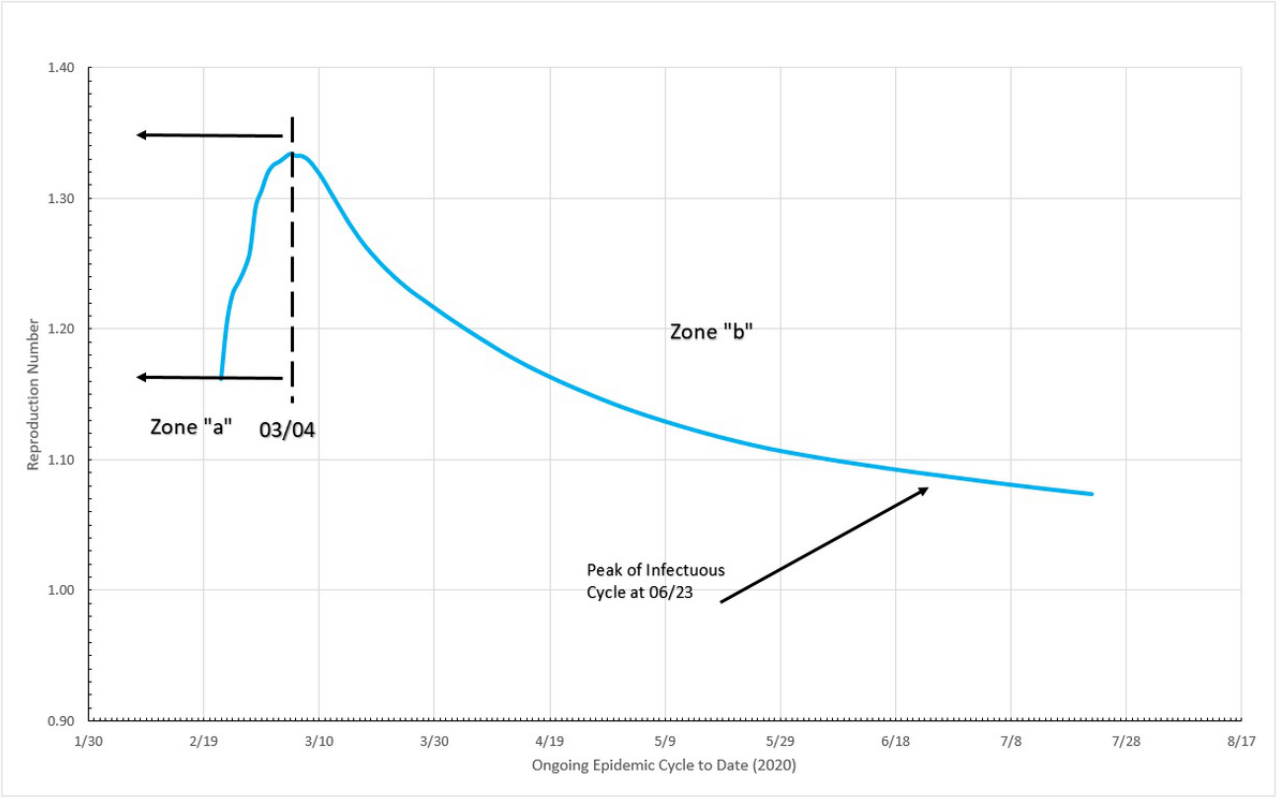


Sweden

Reproduction Numbers

It can be seen in Figure 19 that two distinct zones are formed, once Sweden is considered, by the 5% criteria an “ongoing” epidemic cycle, although in the present date, close to the end. Zone “a” is in the beginning of the cycle, and the reproduction number varies from circa 1.33 to 1.16 from one day to the next; once again this probably is just the reflection of large initial variation in number, but this zone is limited to no more than 5% of the MAMI peak value. It is easy to notice that the figures bear small influence in the overall disease behavior. Zone “b” describes the transmission during the critical disease cycle (from March 4 onward). This is the most lethal period of the epidemic cycle, and it is considered over once a <5% peak level is reached again. In absolute values, the reproduction number for the critical period starts with a value of 1.16 and drops continuously toward 1.07.

Figure 19. Epidemic cycle in Sweden, using the daily number of infected people. Source: Johns Hopkins University [17].



Subnotification

The subnotification effect in Sweden is presented in Figure 20. The calculated limit tells that after the 54th day into the Swedish critical cycle, regardless the amount of subnotification, the error of the calculated reproduction number is no greater than 5%. On the other extreme, a 3x subnotification essentially induces no errors larger than 5% on the reproduction number, after the fourth day during the critical cycle, as shown in Table 9. A maximum error of 18.53% is estimated for the worst case scenario simulated here, a 40x subnotification, and the first day into the cycle. Overall, subnotification appears to have no significant impact in Sweden. Subnotification also has more impact in the very beginning of a given cycle but becomes irrelevant toward the end of it.

Figure 20. Subnotification effect on reproduction number in Sweden during the critical epidemic cycle. Source: Johns Hopkins University [17].

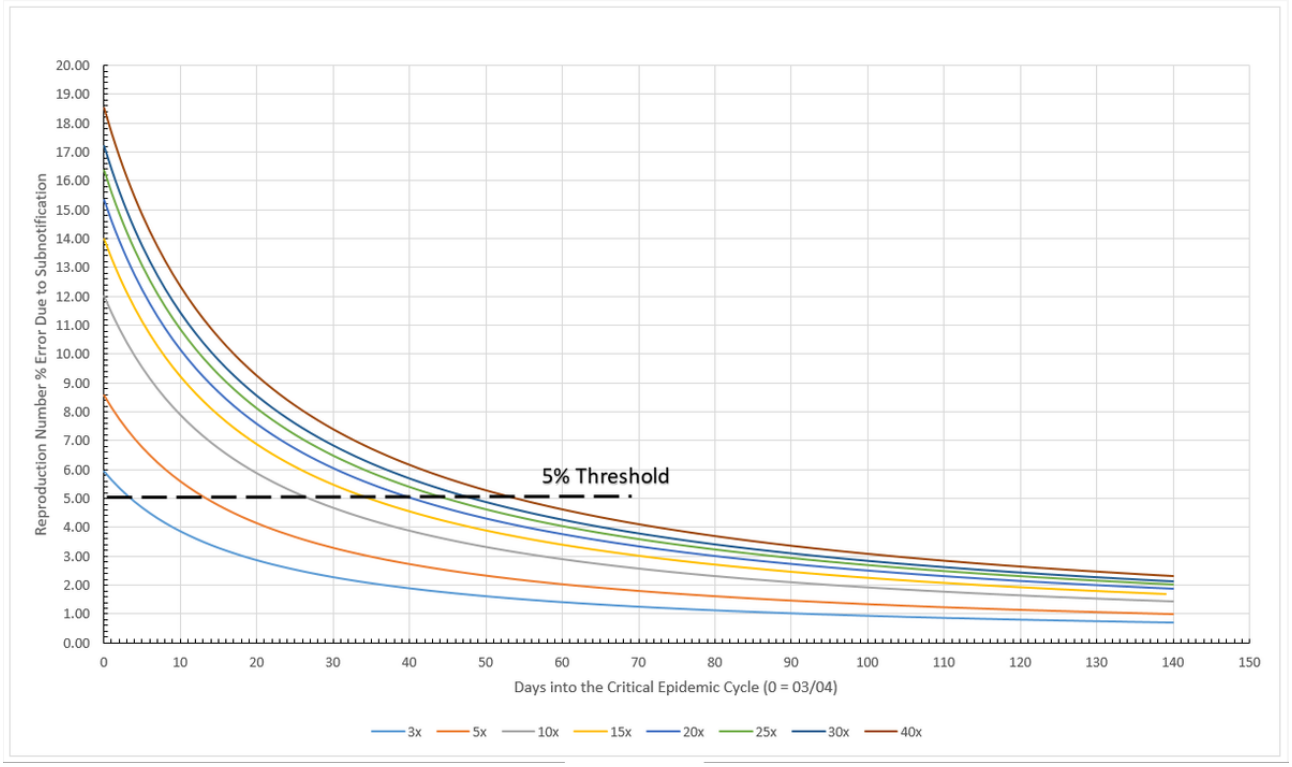


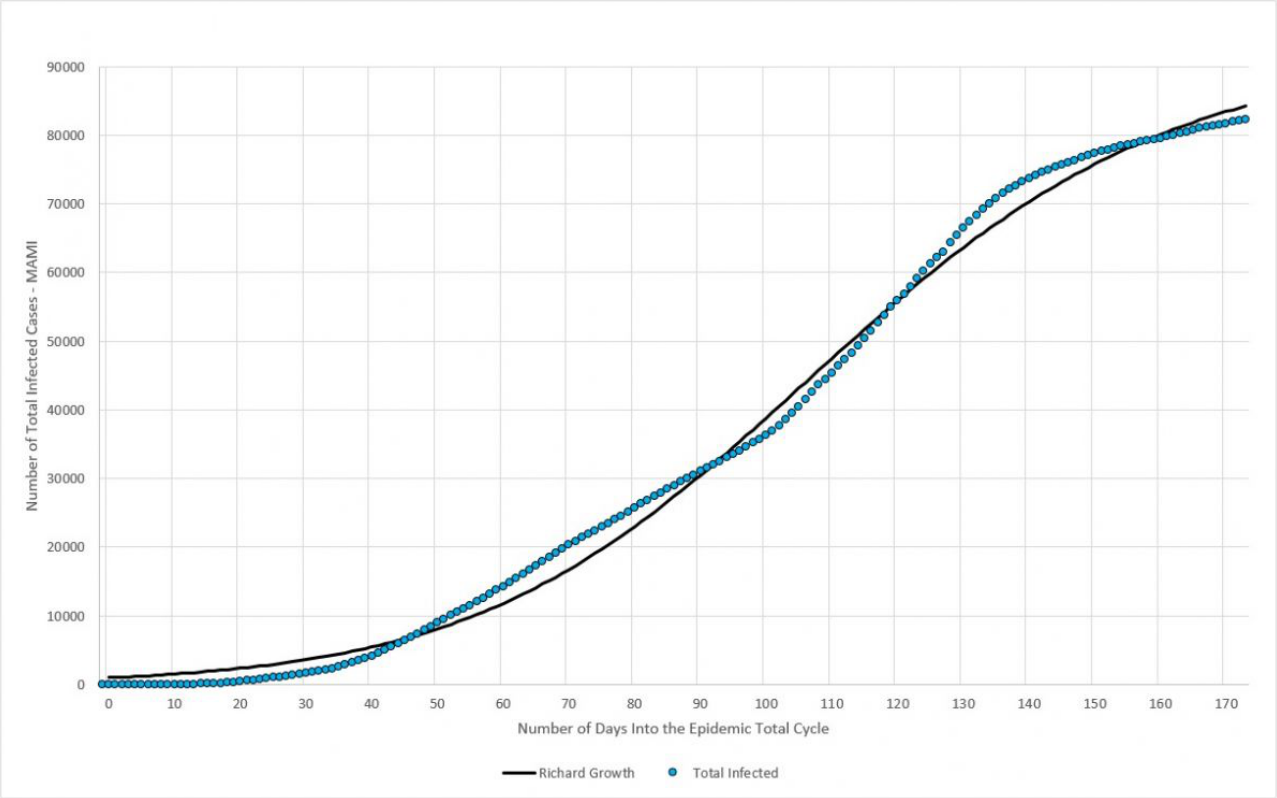
Table 9. Errors associated with ignoring the existence of subnotification in the epidemic cycle for Sweden.

Subnotification	Max error (%)	Min error (%)	Days until ≤5%	Error (%) at peak day
3x	5.92	0.69	4	0.85
5x	8.55	1.01	14	1.24
10x	12.01	1.45	27	1.77
15x	13.97	1.70	35	2.08
20x	15.33	1.88	41	2.30
25x	16.37	2.02	45	2.46
30x	17.22	2.13	49	2.60
40x	18.53	2.31	54	2.82

Total Number of Infected

Data collected for Sweden from February 15 to July 20 were plotted in Figure 21. The blue dots represent the daily registered infected cases submitted to MAMI, and the red continuous line represents the Richard growth model curve, drawn using parameters determined by the MAMI data.

Figure 21. Total number of infected (MAMI [moving average method–initial value]) compared to Richard growth model prediction for Sweden. Source: Johns Hopkins University [17].



Previously, it was shown that the Swedish critical epidemic cycle started on March 4. Using curve-fitting data from Table 10, Table 11 shows that the first case must be recorded 98 days before that, with X_3 indicating that the first case of the total epidemic cycle occurred around November 27, 2019.

Table 10. Curve-fitting data.

Parameter	Value
a	92,538.59
b	3.4050
c	0.0348
d	7.5514×10^{-1}

Table 11. Epidemic parameters determined using curve-fitting data from Table 10.

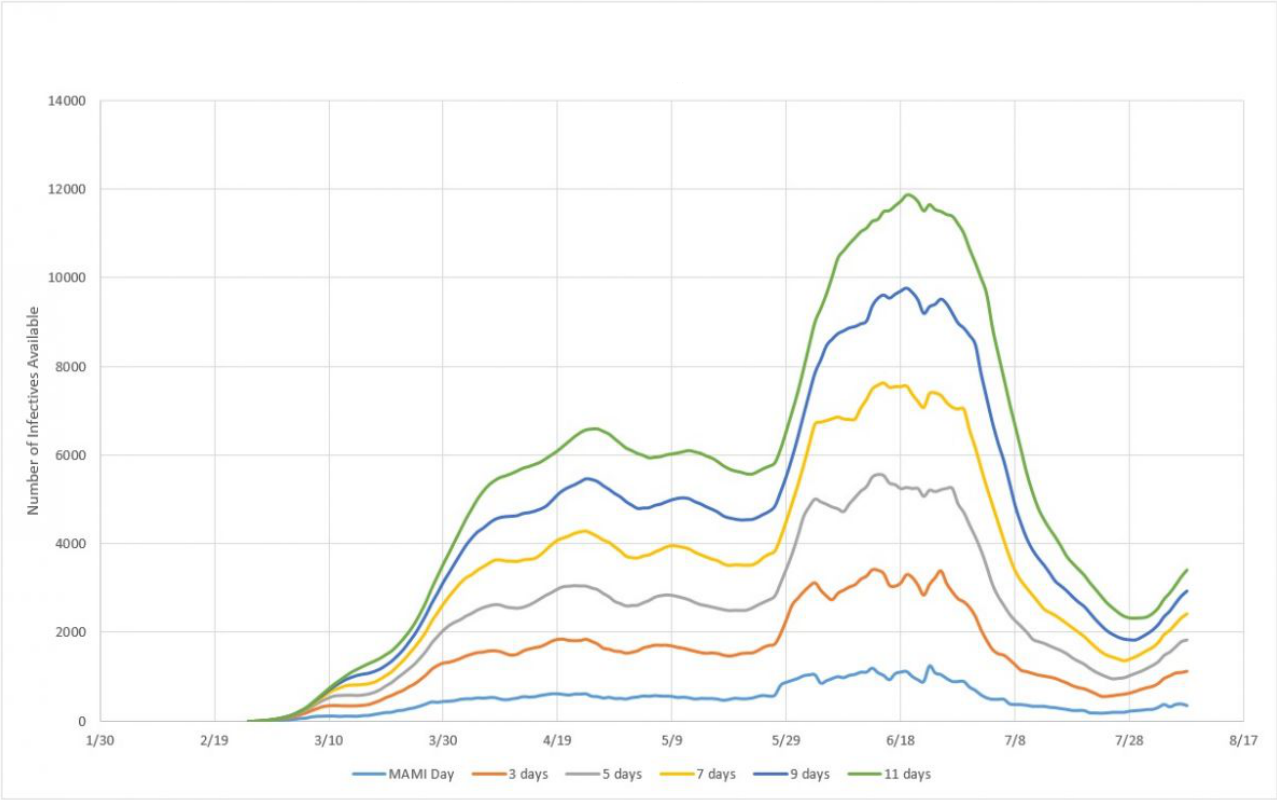
Epidemic parameter	Value
X_1	92,539
X_2	3.48[a]
X_3	98
r^2	0.9958

[a]Percent.

Impact of Incubation Period

Accordingly, Figure 22 presents the predicted inventories of infected persons for Sweden.

Figure 22. Infected person inventories for 3, 5, 7, 9, and 11 days of incubation, compared to MAMI (moving average method–initial value) for Sweden. Source: Johns Hopkins University [17].



One cannot take the assumptions used to derive equation 12 as deterministic, considering that it describes a perfect "production" system. However, there is no biological system that behaves in such a perfect and deterministic way. Therefore, the data shown in Figures 9, 13, and 17 are not conclusive by themselves, given the imperfections of the contamination paths, or the considered "production system," should be taken into account. In other words, the efficiency of the transmission system must be evaluated, as done in the Discussion session.

Discussion

MLCE Control Performance

Using the definition of MLCE, a comparison of the three studied countries was performed. As parameters, it were applied an interval within the 5% limits and the nondimensional time calculated by dividing the day numbers by the total MLCE duration, for each country. For the reproduction number, all the values were divided by the largest value found in the MLCE interval. All these transformations allow us to estimate how efficient the disease control measures used in each country were. In order to enrich the comparative analysis, Figure 23 presents the data from the three countries studied here and also from the United Kingdom, South Korea, and the state of New York. Additional details on this and other comparisons can be found in De Carvalho and De Carvalho [13]. Sweden and New York State were considered as still having an open MLCE by the time of the data analysis; therefore, the end of the cycle considered was the day of data collection (July 22, 2020).

Figure 23. Nondimensional critical epidemic cycle for Germany, Italy, Sweden, South Korea, the United Kingdom, and New York State (NYS). MAMI: moving average method–initial value. Source: Johns Hopkins University [17].

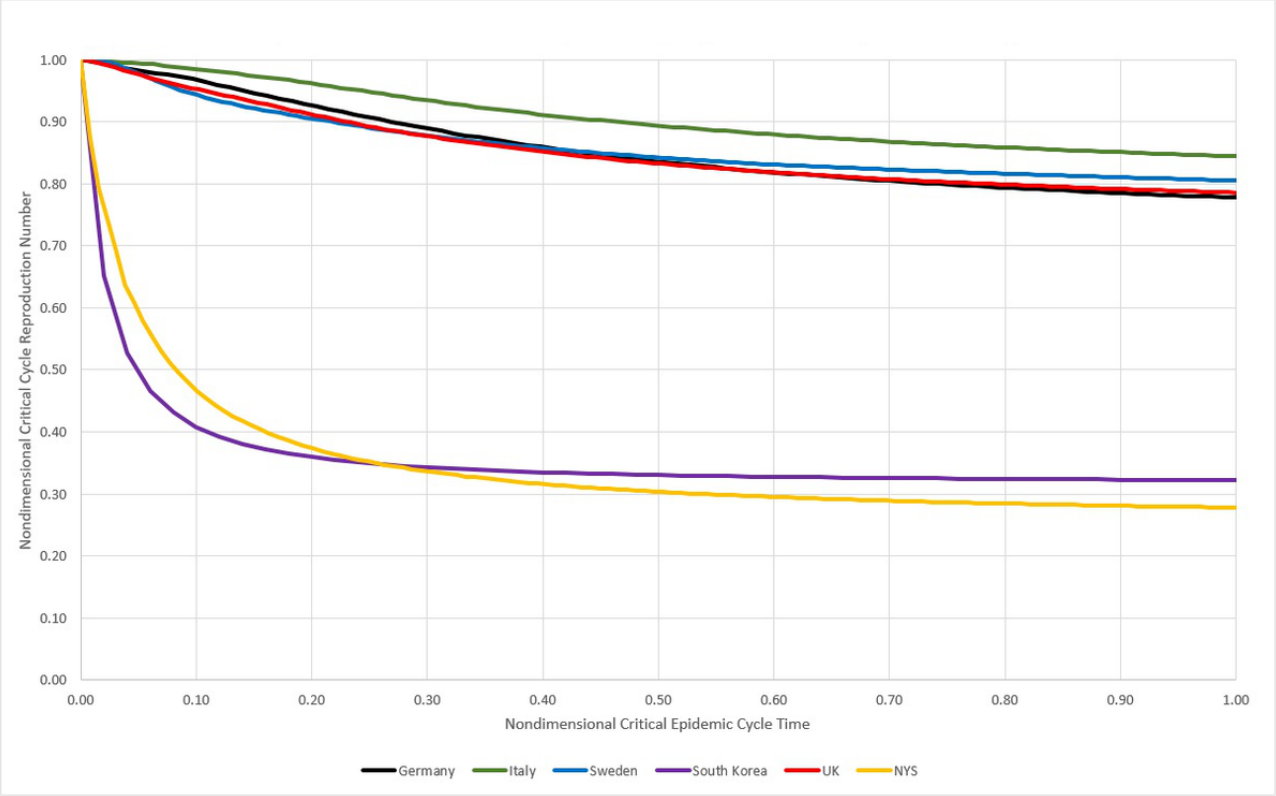
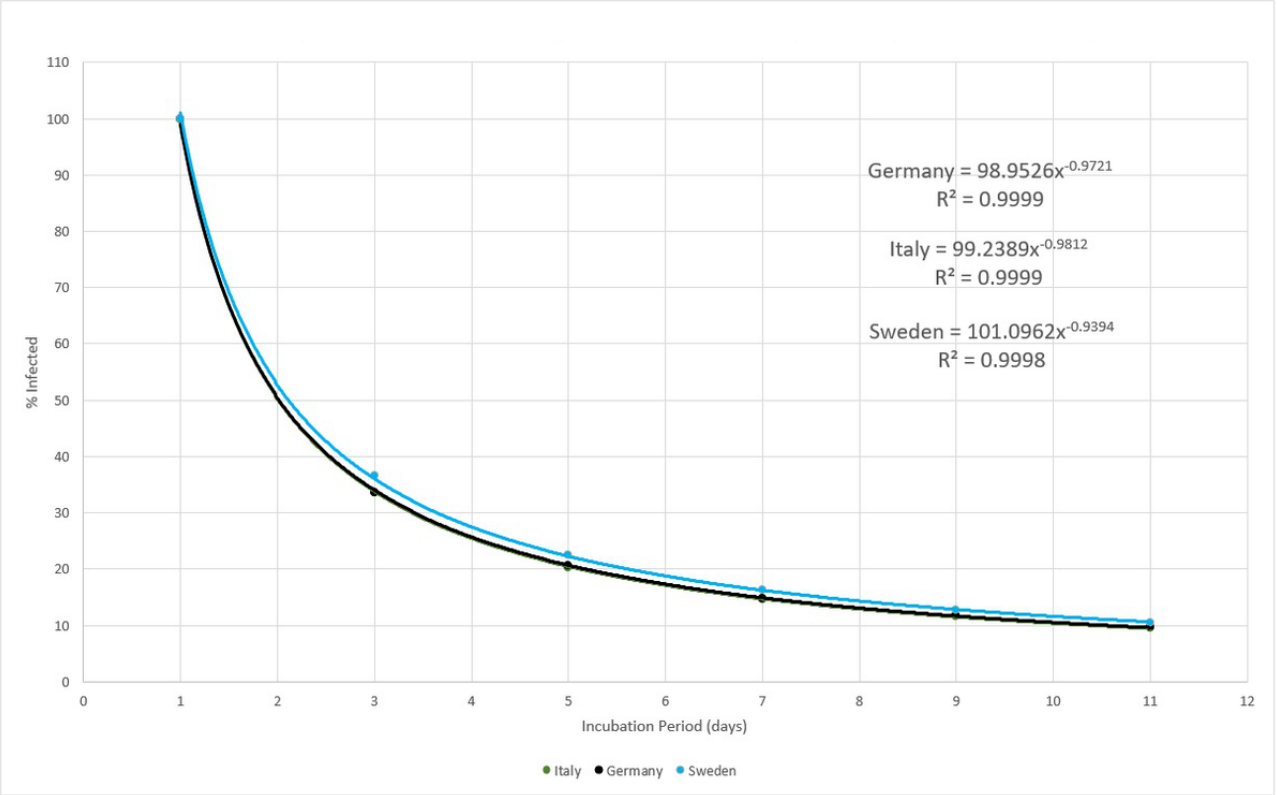


Figure 23 shows that Italy was, in relative terms, the most unsuccessful place in reducing reproduction numbers, although not by a large margin. Germany and the United Kingdom exhibited the same performance where the R_t fell slowly but steadily. South Korea and New York State achieved a large drop in the early stages of the critical cycle, but after that the R_t became more or less constant.

Efficiency of the Infection System

According to the experimental data obtained, the efficiency, or the capacity for spread, of the biological system here described, that is, SARS-CoV-2, has a power function form, as shown in Figure 19. Although the three countries analyzed here present very different epidemic cycles, the percentage of people infected compared to the incubation period varies very little. This probably reflects that the incubation period is in fact a constant value. Figure 24 shows that, for example, for a 5-day incubation period, the percentage of people who were exposed to the virus and displayed symptoms severe enough to prompt them to obtain medical care was around 20%. At the other extreme, if the virus had a 11-day incubation period, the numbers of actual cases registered would have indicated a 10% rate of infection in the general population.

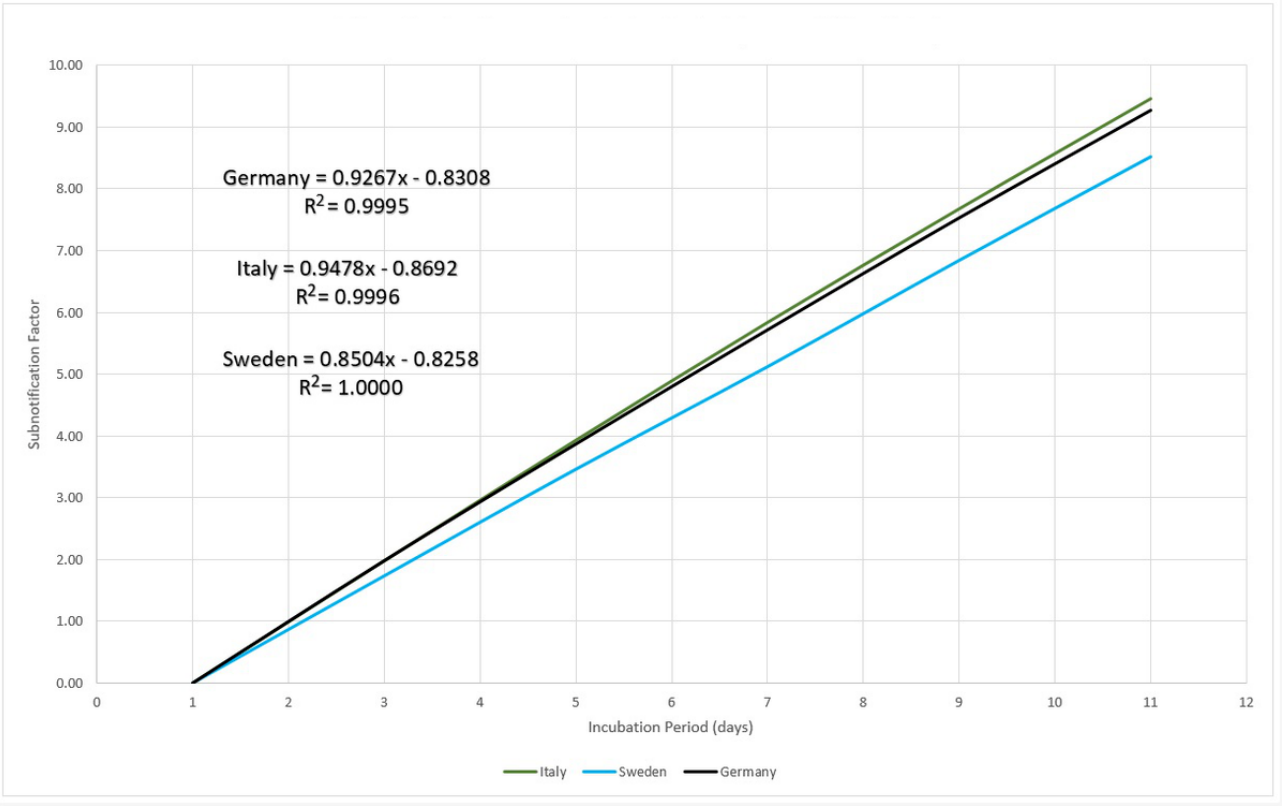
Figure 24. Number of days of incubation versus the percentage of serious and severe COVID-19 infections. Source: Johns Hopkins University [17].



This curve, although restricted to only these three countries, covers nations with quite different NPI policies, population sizes, and land masses. It shows that, according to registered cases, SARS-CoV-2 affected a small segment of these populations and at the same proportions. The subnotification effect does not interfere with this curve behavior significantly, as shown by the calculations.

One conclusion is that, putting together equation 12 with the efficiency measurement in Figure 24, the reported subnotification rate of 80% [22], or 20% of people with more serious symptoms, represents 1 in 5 of the infected persons inventory. In other words, there is 5 times more persons in the infective state than detected and reported by the MAMI figures, leading to a 5-day incubation period. The next step is calculating the subnotification estimation, which then becomes straightforward: given the incubation period, how many times should the registered amount be multiplied to correctly express the estimated subnotification? For example, for a 5-day incubation period in Germany, a subnotification around 4 times the registered number of cases in any given day is expected, if 100 were registered as infected and 400 were not. With this rationale, it is possible to compare the subnotification factor with the incubation period for the three studied countries, as presented in Figure 25.

Figure 25. Subnotification factor for the studied countries. Source: Johns Hopkins University [17].



Other Findings and Conclusions

The early predictions on the progress of the local epidemic cycles of COVID-19 based on Gaussian distribution models and their derivatives, such as the beta distribution, failed to obtain values close to reality, sometimes being very pessimistic, other times being too optimistic. In addition, the nature of the data available for studies requires preliminary numerical treatment, since most of them present the number of daily deaths that occurred on the dates on which they were recorded by the health system and not on those that the deaths actually occurred. Moreover, countries with vast territories and populations should not be treated as a single case, but should be studied regionally, so that the evolution of disease cycles can be clearly understood.

Through the observation of some early cycles, where a peak had already been reached, associated with a consistent reduction in the number of infections, it was possible to identify a triangular shape in these distributions. With the information on the approximate behavior of the variable in question (reproduction number) and the identification of a minimum and maximum, the use of the triangular distribution became clear. After applying this distribution over several local cycles, it was possible to identify similarities between pairs of cycles of localities and regions apparently without direct demographic correlation. Normalization allows you to use an already completed cycle to estimate the behavior of a cycle that is still evolving. The method using the similarity of cycles was able to estimate the end of the cycle up to 34 days before the actual end of the cycle, but requires that there exist a similar cycle. These similarities were confirmed by Kolmogorov-Smirnov tests applied to the data series (Multimedia Appendix 1), demonstrating the hypothesis that the triangular distribution applies to these comparisons and, therefore, is applicable to the prediction of the dimensionless behavior of these cycles. Additionally, understanding the basic behavior of local epidemic cycles allowed for the assessment of the impact of subnotification on calculations.

It is important to note that starting dates influence all the parameters that govern every statistical model used for characterizing the infection. The logistic model together with the model based on the concept of an infected persons inventory can be used to obtain three parameters of the epidemic cycle: the number of total infected, the daily infection rate, and the lag phase, which determines the actual probable onset of the epidemic for the studied countries, thereby solving the problem of noise generation in other parameters by wrongly determined onset dates.

Hence, the experimental framework proposed here offers a set of simple and efficient methods for calculating not only the reproduction number, but also other variables that influence the epidemic cycles and supporting the decision-making process of health authorities, being an interesting tool especially for those places where mass testing is not available. Currently, as the second wave of infections by SARS-CoV-2 emerges, this framework is being applied again in order to definitively demonstrate its efficacy and efficiency.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data and calculations for the studied local cycles.
[XLSX File (Microsoft Excel File), 81 KB - xmed_v2i1e22617_app1.xlsx]

Multimedia Appendix 2
Data and calculations for Rt and subnotification.
[XLSX File (Microsoft Excel File), 383 KB - xmed_v2i1e22617_app2.xlsx]

Multimedia Appendix 3
Extra case study: Brazil.
[DOCX File , 490 KB - xmed_v2i1e22617_app3.docx]

Multimedia Appendix 4
Data and calculations for the logistic model and the inventory model.
[XLSX File (Microsoft Excel File), 226 KB - xmed_v2i1e22617_app4.xlsx]

References

1. Pereira I, Guerin J, Silva Júnior AG, Garcia GS, Piscitelli P, Miani A, et al. Forecasting Covid-19 Dynamics in Brazil: A Data Driven Approach. Int J Environ Res Public Health 2020 Jul 15;17(14):15 [FREE Full text] [doi: 10.3390/ijerph17145115] [Medline: 32679861]
2. Musa S, Zhao S, Wang M, Habib A, Mustapha U, He D. Estimation of exponential growth rate and basic reproduction number of the coronavirus disease 2019 (COVID-19) in Africa. Infect Dis Poverty 2020 Jul 16;9(1):96 [FREE Full text] [doi: 10.1186/s40249-020-00718-y] [Medline: 32678037]
3. Hong HG, Li Y. Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic. PLoS One 2020 Jul 21;15(7):e0236464 [FREE Full text] [doi: 10.1371/journal.pone.0236464] [Medline: 32692753]
4. Kogan NE, Clemente L, Liautaud P. An Early Warning Approach to Monitor COVID-19 Activity with Multiple Digital Traces in Near Real-Time. arXiv. Preprint posted online July 3, 2020 [FREE Full text]
5. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. Nature 2020 Aug 16;584(7821):420-424. [doi: 10.1038/s41586-020-2554-8] [Medline: 32674112]
6. Ogden NH, Fazil A, Arino J, Berthiaume P, Fisman DN, Greer AL, et al. Modelling scenarios of the epidemic of COVID-19 in Canada. Can Commun Dis Rep 2020 Jun 04;46(8):198-204 [FREE Full text] [doi: 10.14745/ccdr.v46i06a08] [Medline: 32673384]
7. Fawad M, Mubarik S, Malik SS, Hao Y, Yu C, Ren J. Trend Dynamics of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission in 16 Cities of Hubei Province, China. CLEP 2020 Jul;12:699-709. [doi: 10.2147/clep.s254806]
8. Zuo M, Khosa SK, Ahmad Z, Almaspoor Z. Comparison of COVID-19 Pandemic Dynamics in Asian Countries with Statistical Modeling. Comput Math Methods Med 2020 Jun 28;2020:4296806-4296816 [FREE Full text] [doi: 10.1155/2020/4296806] [Medline: 32670391]
9. Lin Y, Duan Q, Zhou Y, Yuan T, Li P, Fitzpatrick T, et al. Spread and Impact of COVID-19 in China: A Systematic Review and Synthesis of Predictions From Transmission-Dynamic Models. Front Med (Lausanne) 2020 Jun 18;7:321 [FREE Full text] [doi: 10.3389/fmed.2020.00321] [Medline: 32626719]
10. Chaudhry R, Hanif A, Chaudhary M, Minhas S, Mirza K, Ashraf T, et al. Coronavirus Disease 2019 (COVID-19): Forecast of an Emerging Urgency in Pakistan. Cureus 2020 May 28;12(5):e8346 [FREE Full text] [doi: 10.7759/cureus.8346] [Medline: 32617220]
11. Park S, Bolker B, Champredon D, Earn D, Li M, Weitz J, et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. J R Soc Interface 2020 Jul;17(168):20200144 [FREE Full text] [doi: 10.1098/rsif.2020.0144] [Medline: 32693748]
12. De Carvalho EA, De Carvalho RA. Identification of Patterns in Epidemic Cycles and Methods for Estimating Their Duration: COVID-19 Case Study. JMIR Preprints. Preprint posted July 19, 2020 [FREE Full text] [doi: 10.2196/preprints.22617]
13. De Carvalho EA, De Carvalho RA. COVID-19: Time-Dependent Effective Reproduction Number and Subnotification Effect Estimation Modeling. medRxiv. Preprint posted online August 1, 2020. [doi: 10.1101/2020.07.28.20164087]

14. De Carvalho EA, De Carvalho RA. COVID-19: Estimation of the Actual Onset of Local Epidemic Cycles, Determination of Total Number of Infective, and Duration of the Incubation Period. medRxiv. Preprint posted online August 25, 2020. [doi: 10.1101/2020.08.23.20180356]
15. Worldometer's COVID-19 portal. Worldometer. URL: https://www.worldometers.info/coronavirus/ [accessed 2020-07-09]
16. Kotz S, René van Dorp J. Beyond Beta - Other Continuous Families of Distributions with Bounded Support and Applications. Singapore: World Scientific Publishing Company; 2004.
17. COVID-19 Dashboard. Johns Hopkins University Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-07-22]
18. Hilbe JM. Logistic Regression Models, 1st Edition. Boca Raton, Florida: Chapman & Hall/CRC Texts in Statistical Science; 2017.
19. Bizzarri M, Di Traglia M, Giuliani A, Vestri A, Fedeli V, Prestininzi A. New statistical RI index allow to better track the dynamics of COVID-19 outbreak in Italy. Sci Rep 2020 Dec 22;10(1):22365 [FREE Full text] [doi: 10.1038/s41598-020-79039-x] [Medline: 33353964]
20. Gude-Sampedro F, Fernández-Merino C, Ferreiro L, Lado-Baleato Ó, Espasandín-Domínguez J, Hervada X, et al. Development and validation of a prognostic model based on comorbidities to predict Covid-19 severity. A population-based study. Int J Epidemiol 2020 Dec 08:8 [FREE Full text] [doi: 10.1093/ije/dyaa209] [Medline: 33349845]
21. McCann A, Boice J, Bycoffe A, Silver N, Paine N. 2019-20 NBA Player Projections. FiveThirtyEight. URL: https://projects.fivethirtyeight.com/2020-nba-player-projections/ [accessed 2020-07-09]
22. Similarities and Differences Between COVID-19 and Influenza. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza [accessed 2020-07-23]

Abbreviations

MAM: moving average method
MAMI: moving average method–initial value
MLCE: most lethal cycle of the epidemic
NPI: nonpharmaceutical intervention
R_0: reproduction number
R_t: effective reproduction number

Edited by G Eysenbach, E Meinert; submitted 19.07.20; peer-reviewed by S Facente, M Salman; comments to author 04.11.20; revised version received 18.11.20; accepted 26.12.20; published 18.03.21.

Please cite as:
De Carvalho EA, De Carvalho RA
A Framework for a Statistical Characterization of Epidemic Cycles: COVID-19 Case Study
JMIRx Med 2021;2(1):e22617
URL: https://xmed.jmir.org/2021/1/e22617
doi:10.2196/22617
PMID:34077489

Original Paper

The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study

Carissa Bonner[1], PhD; Erin Cvejic[1], PhD; Julie Ayre[1], PhD; Jennifer Isautier[1]; Christopher Semsarian[2,3,4], PhD; Brooke Nickel[1], PhD; Carys Batcup[1]; Kristen Pickles[1], PhD; Rachael Dodd[1], PhD; Samuel Cornell[1]; Tessa Copp[1], PhD; Kirsten J McCaffery[1], PhD

[1]Sydney Health Literacy Lab, School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Agnes Ginges Centre for Molecular Cardiology, Centenary Institute, The University of Sydney, Sydney, Australia

[3]Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[4]Department of Cardiology, Royal Prince Alfred Hospital, Sydney, Australia

Corresponding Author:
Carissa Bonner, PhD
Sydney Health Literacy Lab, School of Public Health
Faculty of Medicine and Health
The University of Sydney
Edward Ford Building A27, Rm 128A
Sydney, NSW 2006
Australia
Phone: 61 2 9351 7125
Email: carissa.bonner@sydney.edu.au

Related Articles:

Companion article: https://preprints.jmir.org/preprint/25610

Companion article: https://med.jmirx.org/2021/1/e28714/

Companion article: https://med.jmirx.org/2021/1/e28717/

Companion article: https://med.jmirx.org/2021/1/e28718/

Abstract

Background: It is unclear how people with hypertension are responding to the COVID-19 pandemic given their increased risk, and whether targeted public health strategies are needed.

Objective: This retrospective case-control study compared people with hypertension to matched healthy controls during the COVID-19 lockdown to determine whether they have higher risk perceptions, anxiety, and vaccination intentions.

Methods: Baseline data from a national survey were collected in April 2020 during the COVID-19 lockdown in Australia. People who reported hypertension with no other chronic conditions were randomly matched to healthy controls of similar age, gender, education, and health literacy level. A subset including participants with hypertension was followed up at 2 months after restrictions were eased. Risk perceptions, anxiety, and vaccination intentions were measured in April and June.

Results: Of the 4362 baseline participants, 466 (10.7%) reported hypertension with no other chronic conditions. A subset of 1369 people were followed up at 2 months, which included 147 (10.7%) participants with hypertension. At baseline, perceived seriousness was high for both hypertension and control groups. The hypertension group reported greater anxiety compared to the controls and were more willing to vaccinate against influenza, but COVID-19 vaccination intentions were similar. At follow-up, these differences were no longer present in the longitudinal subsample. Perceived seriousness and anxiety had decreased, but vaccination intentions for both influenza and COVID-19 remained high across groups (>80%).

Conclusions: Anxiety was above normal levels during the COVID-19 lockdown. It was higher in the hypertension group, which also had higher vaccination intentions. Groups that are more vulnerable to COVID-19 may require targeted mental health screening

during periods of greater risk. Despite a decrease in perceived risk and anxiety after 2 months of lockdown restrictions, vaccination intentions remained high, which is encouraging for the future prevention of COVID-19.

(JMIRx Med 2021;2(1):e25610) doi:10.2196/25610

KEYWORDS

public health; global health; COVID-19; hypertension; risk; strategy; mental health; behavior; response; anxiety; vaccine; retrospective; perception; prevention; intention

Introduction

Although research on COVID-19 outcomes is constantly evolving, there is consistent evidence that people with cardiovascular disease (CVD) risk factors are more likely to experience severe complications and are more likely to die if they acquire COVID-19 [1]. People with CVD are more likely to have risk factors that may complicate the response to COVID-19, and COVID-19 can itself cause cardiovascular damage [2]. During the early phase of the pandemic, there was prominent media attention about the risk of hypertension in particular, and there were concerns that people with CVD risk factors were not presenting to general practitioners and hospitals for management and new symptoms onset due to the fear of contracting COVID-19 [3,4]. People with CVD risk factors or established CVD can access prescriptions via telehealth in Australia, but this was very new at the time of the study [5]. As well as potential access issues, many people with chronic conditions do not believe they are at increased risk, which may affect their uptake of prevention measures [6]. This may be reinforced by beliefs based on misinformation about the severity of COVID-19, spread as part of antivaccination movements [7].

In addition to concern about increased risk for this population, there has been debate in the medical community about whether common medications used to manage risk for people with CVD, hypertension, and diabetes contribute to worse COVID-19 outcomes [8,9]. At the time of this study, there was insufficient evidence to cease their use, prompting the National Heart Foundation to release a statement confirming this [10]. However, there continues to be research on the role of angiotensin converting enzyme inhibitors and angiotensin II type I receptor blockers, with arguments both for and against the continued use of such medications [11,12] during the COVID-19 pandemic in different population groups.

There has also been debate about the respiratory versus cardiovascular nature of COVID-19. Emerging research suggests that virus complications and their treatment could be regarded as cardiovascular in nature [13,14], which may explain the devastating outcomes experienced by some people who contract the virus. It is unclear what this means for managing people with multiple CVD risk factors associated with worse COVID-19 outcomes (eg, hypertension and diabetes) [8]. Initial concerns promoted in national media included both respiratory conditions, such as asthma [15], and cardiovascular conditions, including hypertension [16], early in the Australian pandemic response.

As a result of this evolving and conflicting research, as well as widespread misinformation, people with hypertension in the community may have received mixed messages in the media about how they should manage both CVD risk and COVID-19 risk during the pandemic. It is unknown whether people with hypertension responded differently to the pandemic and associated restrictions compared to the general population and whether a tailored communication approach is needed to address the needs of this group.

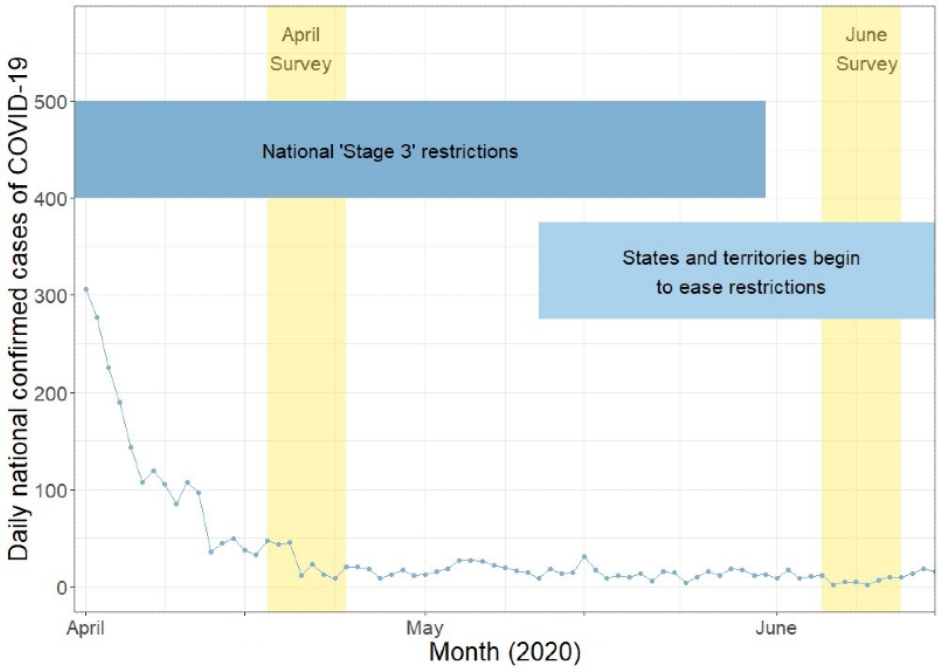
This study investigated whether people with hypertension have higher risk perceptions, anxiety, and prevention intentions during COVID-19 restrictions to inform targeted public health messaging for this group.

Methods

Setting

In Australia, the COVID-19 pandemic has been well controlled compared to many other countries around the world. However, in April 2020, cases and community transmissions had been rising exponentially, and the country was placed under lockdown, including closure of schools and workplaces and restrictions on gatherings and movement. Citizens were required to stay home except for essential purposes (eg, work, essential shopping, exercise). In June 2020, cases were under control and many regulations were eased, although some restrictions remained, such as small gathering sizes, which varied from state to state. A second wave occurred in the state of Victoria shortly after this, requiring new restrictions such as mandatory masks and curfews, but our data were collected prior to this. Thus, a comparison of April and June data presents an opportunity to look at the effect of a short-term lockdown between a time of strong COVID-19 restrictions and good control (Figure 1).

Figure 1. COVID-19 in Australia during the study period.



Data Collection

Data from a national Australian survey were used to conduct retrospective case-control analyses comparing hypertension and control groups. Baseline data were collected from all states and territories in April 2020 during the COVID-19 lockdown, with a subsample followed up in June 2020 when restrictions were eased.

Ethics approval was obtained from the University of Sydney Human Research Ethics Committee (2020/212).

Measures

The survey measures and full sample results are reported elsewhere [17,18], including the Health Literacy single-item screener [19], Consumer Health Activation Index (CHAI) patient activation measure [20], and State-Trait Anxiety Inventory (STAI) [21]. Participants were asked if they had any of the following conditions: asthma, chronic obstructive pulmonary disease, high blood pressure (hypertension), cancer, heart disease, stroke, diabetes, depression, or anxiety; and whether they take any prescription medication (not specified). The single-item screener provides a brief measure of health literacy, that is, the skills needed to engage in health [19]; and the CHAI provides a measure of patient activation, that is, the extent that a person actively involves themselves in decisions to manage one's health [20]. Risk perceptions and prevention behaviors (including vaccination intentions) were measured using Likert and categorical scales. Items pertaining to risk perception were based on items developed for an earlier US COVID-19 study [18]. The perceived seriousness of threat from COVID-19 was captured using a 10-point scale (1="no threat at all to" 10="very serious public health threat"). The social distancing score reflects perceived importance of social distancing. This outcome is based on 4 items, each answered using a 7-point Likert scale. The items were adapted from existing vaccine attitude instruments to instead reflect on social distancing ("social distancing is important for my family's health," "social distancing is important for the health of others in my community," "when everyone else is socially distancing, I don't need to," "I socially distance to protect people with a weaker immune system"). Perceived seriousness was asked generally at baseline; at follow-up, participants were asked about the public health risk from COVID-19 in general, globally, and in Australia specifically, given the divergent pattern of control across countries.

Matching Procedure

Individuals with hypertension and no other comorbidities (n=466) were retrospectively matched without replacement to healthy controls (with no comorbidities; n=2251) using the *calipmatch* function in Stata (StataCorp) [22]. For each case, potential controls were initially identified based on age (±3 years) and exactly matching on gender, education, and health literacy adequacy (selected given observed differences as a function of these variables in COVID-19–related knowledge, attitudes, awareness, and behaviors in our baseline survey [17]). One matching control is then randomly selected for the case and removed from the list of available controls for subsequent cases. Because the search strategy for controls is greedy (ie, selecting cases for matching in random order and removing controls without replacement for subsequent case matching), some cases may be left unmatched. The initial matching run resulted in 95.7% (446/466) of cases successfully matched to a control. The constraints for matching were iteratively relaxed (eg, allowing age to vary by ±10 years; education level to differ by one category) until all remaining cases were paired to a control. The matching procedure was repeated for the follow-up sample.

Analysis

Analyses were conducted using Stata/IC v16.1 (StataCorp). Pairwise comparisons of baseline demographic characteristics were undertaken to confirm the appropriateness of the matching

procedure of cases to controls, and to identify potential differences in demographic characteristics between those who were invited and returned for follow-up compared to those who were not followed up. Regression models with robust error variances to account for clustering within pairs, and adjusted for matching variables (age, gender, education, and health literacy adequacy), were used to analyze outcome variables. Linear models were used for continuous outcomes (risk perceptions, STAI anxiety, perceived importance of social distancing) to estimate marginal mean differences (MMD). Generalized linear models with a modified Poisson approach [23] were used for the dichotomous outcome "not feeling stressed due to COVID-19," generating adjusted prevalence ratios (aPR). Ordinal logistic regression models were used for ordered categorical outcomes (frequency of leaving one's home, vaccination intentions), resulting in adjusted odds ratios (aOR). Separate models were conducted for each time point. All estimates are provided with 95% CI values. A *P* value of .05 was used as the threshold for statistical significance.

Data Availability

Data are available upon reasonable request subject to ethics approval.

Results

Of the 4362 baseline participants, 466 (10.7%) reported hypertension with no other chronic conditions. A subset of 1369 participants from the original survey cohort were followed up after 2 months, comprising 147 (10.7%) participants with hypertension only.

Table 1 describes the case versus control samples for all baseline outcomes, and Table 2 shows details of the regression models comparing the two groups at this timepoint. Table 3 provides a description of cases and controls included in the follow-up sample, with Table 4 detailing the outcome of the regression models at follow-up.

Table 1. Baseline descriptive statistics and unadjusted outcomes for hypertension cases versus matched healthy controls.

Variable	Group[a]	
	Hypertension (n=466)	Control (n=466)
Sample description		
Age (years), mean (SD)	53.5 (15.5)	52.5 (15.3)
Age group, n (%)		
18-25 years	26 (6)	34 (7)
26-40 years	83 (18)	78 (17)
41-55 years	105 (23)	117 (25)
56-90 years	252 (54)	237 (51)
Gender, n (%)		
Male	220 (47)	220 (47)
Female	243 (52)	243 (52)
Not specified/other	3 (1)	3 (1)
Education, n (%)		
High school or less	115 (25)	112 (24)
Certificate I-IV	69 (15)	69 (15)
University	282 (61)	285 (61)
Adequate health literacy[b], n (%)	427 (92)	431 (92)
Takes any prescription medicine, n (%)	359 (77)	195 (42)
Consumer Health Activation Index (score 0-100 where 100 is more active), mean (SD)	75.83 (14.19)	77.17 (12.77)
Risk perception		
Seriousness of threat (0=low, 10=high), mean (SD)	7.72 (2.25)	7.66 (2.18)
What percentage of people who get COVID-19 will die as a result? (open), mean (SD)	6.50 (13.49)	5.72 (12.45)
What percentage of people who get COVID-19 will experience only mild symptoms? (open), mean (SD)	62.88 (26.36)	62.37 (27.12)
Anxiety		
State-Trait Anxiety Inventory (score range 20-80; normal 34-36), mean (SD)	40.62 (14.95)	38.98 (14.38)
Never (in the past week) felt nervous or stressed because of COVID-19 (categorical), n (%)	113 (24)	115 (25)
Prevention behaviors		
Perceived importance of social distancing (average of 4 items from 1-7, where 7 is most important), mean (SD)	6.48 (0.74)	6.42 (0.82)
How often are you leaving home? n (%)		
Less than once per week	45 (10)	42 (9)
Once per week	53 (11)	53 (11)
A few times per week	176 (38)	150 (32)
Once per day	154 (33)	176 (38)
Multiple times per day	38 (8)	45 (10)
I have or I will get the flu vaccine this year, n (%)		
Strongly disagree/disagree	50 (11)	72 (15)
Neither agree nor disagree	30 (6)	39 (8)
Strongly agree/agree	386 (83)	355 (76)

Variable	Group[a]	
	Hypertension (n=466)	Control (n=466)
If a COVID-19 vaccine becomes available, I will get it, n (%)		
Strongly disagree/disagree	17 (4)	29 (6)
Neither agree nor disagree	45 (10)	42 (9)
Strongly agree/agree	404 (87)	395 (85)

[a]People reporting high blood pressure and no other conditions were matched to healthy controls with no reported cardiovascular or respiratory conditions.
[b]Based on the single-item health literacy screener.

Table 2. Multivariable[a] regression model estimates comparing hypertension cases (n=466) versus matched healthy controls (n=466) at baseline.

Variable	Estimate (95% CI)	*P* value
Risk perception		
Seriousness of threat, MMD[b]	0.05 (–0.23 to 0.34)	.71
What percentage of people who get COVID-19 will die as a result? MMD	0.75 (–0.87 to 2.37)	.36
What percentage of people who get COVID-19 will experience only mild symptoms? MMD	0.71 (–2.77 to 4.18)	.69
Anxiety		
State-Trait Anxiety Inventory, MMD	1.90 (0.19 to 3.61)	.03
Never (in the past week) felt nervous or stressed because of COVID-19, aPR[c]	0.96 (0.77 to 1.19)	.69
Prevention behaviors		
Perceived importance of social distancing, MMD	0.06 (–0.04 to 0.17)	.21
How often are you leaving home? aOR[d]	0.84 (0.66 to 1.06)	.14
I have or I will get the flu vaccine this year, aOR	1.52 (1.10 to 2.11)	.01
If a COVID-19 vaccine becomes available, I will get it, aOR	1.21 (0.84 to 1.73)	.31

[a]All multivariable models controlled for age (in years), gender, health literacy adequacy, and education.
[b]MMD: marginal mean difference (from the linear regression model).
[c]aPR: adjusted prevalence ratio (from the generalized linear model using a modified Poisson approach).
[d]aOR: adjusted odds ratio (from the ordinal logistic regression).

Table 3. Follow-up descriptive statistics and unadjusted outcomes for hypertension cases versus matched healthy controls.

Variable	Group[a]	
	Hypertension (n=147)	Control (n=147)
Sample description[b]		
Age (years), mean (SD)	54.8 (14.9)	52.8 (14.2)
Age group, n (%)		
18-25 years	7 (5)	8 (5)
26-40 years	22 (15)	22 (15)
41-55 years	36 (24)	45 (31)
56-90 years	82 (56)	72 (49)
Gender, n (%)		
Male	61 (41)	61 (41)
Female	85 (58)	85 (58)
Not specified/other	1 (1)	1 (1)
Education, n (%)		
High school or less	26 (18)	18 (12)
Certificate I-IV	19 (13)	21 (14)
University	102 (69)	108 (73)
Adequate health literacy[c], n (%)	142 (97)	143 (97)
Takes any prescription medicine, n (%)	114 (78)	56 (38)
Consumer Health Activation Index (score 0-100, where 100 is more active), mean (SD)	75.48 (14.32)	77.10 (12.95)
Risk perception		
Seriousness of threat in general (0=low to 10=high), mean (SD)	7.51 (2.42)	7.03 (2.58)
Seriousness of threat globally (0=low to 10=high), mean (SD)	8.74 (1.76)	8.65 (1.81)
Seriousness of threat in Australia (0=low to 10=high), mean (SD)	6.14 (2.38)	5.50 (2.49)
Anxiety		
State-Trait Anxiety Inventory (score range 20-80; normal 34-36), mean (SD)	36.94 (15.31)	36.49 (13.93)
Never (in the past week) felt nervous or stressed because of COVID-19 (categorical), n (%)	58 (39)	64 (44)
Prevention behaviors		
Perceived importance of social distancing (average of 4 items from 1-7, where 7 is more important), mean (SD)	6.49 (0.78)	6.34 (0.90)
I have or I will get the flu vaccine this year, n (%)		
Strongly disagree/disagree	13 (9)	24 (16)
Neither agree nor disagree	2 (1)	2 (1)
Strongly agree/agree	132 (90)	121 (82)
If a COVID-19 vaccine becomes available, I will get it, n (%)		
Strongly disagree/disagree	7 (5)	13 (9)
Neither agree nor disagree	9 (6)	10 (7)
Strongly agree/agree	131 (89)	124 (84)

[a]People reporting high blood pressure and no other conditions were matched to healthy controls with no reported cardiovascular or respiratory conditions.

[b]As measured at baseline.

[c]Based on the single-item health literacy screener.

Table 4. Multivariable[a] regression model estimates comparing hypertension cases (n=147) versus matched healthy controls (n=147) at follow-up.

Variable	Estimate (95% CI)	*P* value
Risk perception		
Seriousness of threat in general, MMD[b]	0.50 (–0.08 to 1.08)	.09
Seriousness of threat globally, MMD	0.07 (–0.31 to 0.46)	.71
Seriousness of threat in Australia, MMD	0.60 (0.05 to 1.15)	.03
Anxiety		
State-Trait Anxiety Inventory, MMD	0.94 (–2.57 to 4.45)	.60
Never (in the past week) felt nervous or stressed because of COVID-19, aPR[c]	1.03 (0.94 to 1.12)	.55
Prevention behaviors		
Perceived importance of social distancing, MMD	0.16 (–0.03 to 0.35)	.11
I have or I will get the flu vaccine this year, aOR[d]	1.90 (0.93 to 3.90)	.08
If a COVID-19 vaccine becomes available, I will get it, aOR	1.72 (0.82 to 3.58)	.15

[a]All multivariable models controlled for age (in years), gender, health literacy adequacy, and education.

[b]MMD: marginal mean difference (from the linear regression model).

[c]aPR: adjusted prevalence ratio (from the generalized linear model using a modified Poisson approach).

[d]aOR: adjusted odds ratio (from the ordinal logistic regression).

Description of Sample

To isolate the effects of hypertension, the hypertension sample included 466 people reporting only high blood pressure and no other chronic health conditions. The mean age was 54 years (SD 15.5), and the sample comprised 52% (n=243) female, 47% (n=220) male, and 1% (n=3) unspecified. The majority had a university degree (n=282, 61%) and adequate health literacy (n=427, 92%). The average patient activation score was comparable to other patient populations (mean scaled CHAI 74.9). Most were taking medications (n=359, 77%), with 45% (n=163) obtaining a refill during the lockdown, 5% (n=19) switching to a longer prescription, and only 1 person stopping their medication. As seen in Table 1, the sample descriptive characteristics were comparable between individuals with hypertension and the matched controls. There was no statistical difference across age (*P*=.33), gender (*P*>.99), education (*P*=.97), or health literacy adequacy (*P*=.63) between cases and controls. Cases who were invited and returned for follow-up were of similar age and gender but had higher levels of education (*P*=.02) and were more likely to have adequate health literacy (*P*=.009) than those who were not followed up.

Risk Perceptions

At baseline, the perceived seriousness of threat from COVID-19 in the hypertension group was high (mean 7.72, out of 10) but similar to controls (mean 7.66). On average, the hypertension sample believed that 7% of people who get COVID-19 would die as a result and 63% would experience only mild symptoms (asked separately). There were no statistically significant differences between the hypertension group and the matched controls at baseline. At follow-up, those with hypertension perceived a greater threat (mean 6.12) than controls (mean 5.52) when asked about Australia (MMD 0.60, 95% CI 0.05-1.15; *P*=.03) but not in general or globally.

Anxiety

At baseline, 76% (n=353) of the hypertension group had felt nervous or stressed about COVID-19 in the past week at least some of the time. On average, the mean STAI was 1.90 units higher (95% CI 0.19 to 3.61; *P*=.03, Cohen *d*=0.13) for those with hypertension (mean 40.75) than matched controls (mean 38.85). At follow-up, there was no longer a significant difference between the hypertension (mean 37.02) and control (mean 36.08) groups (MMD 0.94, 95% CI –2.57 to 4.45; *P*=.60, Cohen *d*=0.06).

Prevention Behaviors

At baseline, the hypertension group had a social distancing score of 6.48 out of 7, indicating strong agreement with the importance of social distancing for ones' own health and the health of the public; this was similar to the controls (6.42 out of 7). Most people left home a few times a week (n=176, 38%) or once a day (n=154, 33%) during the lockdown. Overall, 83% (n=386) agreed they would get the influenza vaccine, and 87% (n=404) would get the COVID-19 vaccine. Compared to healthy matched controls, the hypertension group was more likely to agree that they would (or have already) received the influenza vaccine this year (aOR 1.52, 95% CI 1.10-2.11; *P*=.01). There were no significant differences in willingness to vaccinate for COVID-19 (if it became available), perceived importance of social distancing, or frequency of leaving one's home. At follow-up, there was no longer a significant difference between the hypertension and control groups for influenza vaccination intention (aOR 1.90, 95% CI 0.93-3.90; *P*=.08), with intentions remaining high for both influenza and COVID-19 vaccination (>80% for both groups).

Discussion

Principal Findings

The main observation of this study was the significant difference in anxiety levels between hypertension only and matched control groups, with all groups reporting higher than "normal" levels. This is consistent with the Australian Bureau of Statistics' finding that the rate of anxiety in the general population had doubled in April 2020 compared to a survey from 2017-18 [24]. Prioritizing mental health screening for more vulnerable clinical groups with higher anxiety may be warranted when local community transmission rates are high.

Overall, there were few differences between people with hypertension and healthy matched controls. No significant differences were found for COVID-19 risk perceptions or perceived importance of social distancing behaviors. This is consistent with another study, which found that 20% of people with chronic conditions did not perceive greater risk [6], but differs from other survey reports that indicate people with different chronic conditions are more likely to engage in COVID-19 prevention behaviors and perceive COVID-19 as a serious threat [18,25]. This may be due to a close resemblance between the hypertension and general populations in our study, or it may be a result of our method of matching cases to controls rather than comparing groups without such adjustment. Another Australian survey found similarly high risk perceptions, so there may also be a ceiling effect in Australia across community groups [26].

Responses to flu vaccine uptake varied across the two groups, whereby those with hypertension were more likely to intend to vaccinate compared to healthy controls. It is possible this is due to the former's greater exposure to the health system where doctors may mention the flu vaccine each year. This difference does not appear to transfer to increased intent for COVID-19 vaccine uptake, but this may be due to a ceiling effect with high acceptance rates in Australia [27] compared to other countries such as France [28]. It should be noted that vaccine acceptance rates are changing over time as new information (and misinformation) becomes available about the various vaccines [29] now being used around the world. No COVID-19 vaccinations were available to Australians at the time of the study in 2020.

Differences in medication use were found between groups, but this was to be expected given that preventive medication is recommended for hypertension. Surprisingly, there were no differences in access difficulties or changes to medication. The Australian Bureau of Statistics reported in April 2020 that almost half (47%) of respondents with a chronic condition had used telehealth [24], including electronic prescriptions; this was not a focus of our survey but may explain why little change was detected.

Strengths and Limitations

The strengths of this study include a large national sample with data during and after lockdown restrictions, which enabled matched case-control analyses between participants with self-reported hypertension and healthy controls and the use of established, well-validated measures.

The sample was recruited via an online research panel and social media, and has a low proportion of culturally and linguistically diverse participants; hence, different results may be found in other populations. We are currently conducting a separate survey of these communities in their preferred language. The survey involved nonstratified sampling without targeted recruitment of specific health conditions, and only a subset were included in the longitudinal substudy. Future research could explore the influence of multimorbidity and differences between social media users and other community members, given misinformation concerns in Australia [30].

Conclusion

Anxiety was above normal levels for all groups during the COVID-19 lockdown. This was higher among people with hypertension, who also had higher influenza vaccination intentions but similar COVID-19 vaccination intentions. In Australia, where lockdown measures effectively reduced the spread of COVID-19 and restrictions eased relatively quickly, these differences dissipated after 2 months, but locations with prolonged restrictions may require targeted psychological screening for vulnerable groups. Despite a decrease in perceived seriousness and anxiety after 2 months of lockdown restrictions, vaccination intentions for both influenza and COVID-19 remained high (80%), which is encouraging for the future prevention of COVID-19.

Acknowledgments

This study received no direct funding, but in-kind support was provided by authors with research fellowships. The SHeLL (Sydney Health Literacy Lab) group thanks the participants of the longitudinal COVID-19 survey for their participation in this research. CB is supported by a National Health and Medical Research Council (NHMRC)/Heart Foundation Early Career Fellowship (1122788). CS is the recipient of a NHMRC Practitioner Fellowship (1154992). RD is supported by a University of Sydney fellowship (197589). KM is supported by an NHMRC Principal Research Fellowship (1121110).

Conflicts of Interest

None declared.

References

1. Zaman S, MacIsaac AI, Jennings GL, Schlaich MP, Inglis SC, Arnold R, et al. Cardiovascular disease and COVID-19: Australian and New Zealand consensus statement. Med J Aust 2020 Aug 31;213(4):182-187. [doi: 10.5694/mja2.50714] [Medline: 32734645]
2. Li G, Hu R, Gu X. A close-up on COVID-19 and cardiovascular diseases. Nutr Metab Cardiovasc Dis 2020 Jun 25;30(7):1057-1060 [FREE Full text] [doi: 10.1016/j.numecd.2020.04.001] [Medline: 32423664]
3. Thornton J. Covid-19: A&E visits in England fall by 25% in week after lockdown. BMJ 2020 Apr 06;369:m1401. [doi: 10.1136/bmj.m1401] [Medline: 32253175]
4. Kippen R, O'Sullivan B, Hickson H, Leach M, Wallace G. A national survey of COVID-19 challenges, responses and effects in Australian general practice. Aust J Gen Pract 2020 Nov 01;49(11):745-751. [doi: 10.31128/ajgp-06-20-5465]
5. Ministers | Department of Health. COVID-19: Whole of population telehealth for patients, general practice, primary care and other medical services. Commonwealth of Australia. 2020 Mar 29. URL: https://tinyurl.com/tzh8u29s [accessed 2021-02-16]
6. Tran V, Ravaud P. COVID-19-related perceptions, context and attitudes of adults with chronic conditions: Results from a cross-sectional survey nested in the ComPaRe e-cohort. PLoS One 2020 Aug 6;15(8):e0237296 [FREE Full text] [doi: 10.1371/journal.pone.0237296] [Medline: 32760127]
7. Vrdelja M, Kraigher A, Vercic D, Kropivnik S. The growing vaccine hesitancy: exploring the influence of the internet. Eur J Public Health 2018 Oct 01;28(5):934-939. [doi: 10.1093/eurpub/cky114] [Medline: 29982349]
8. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? The Lancet Respiratory Medicine 2020 Apr;8(4):e21. [doi: 10.1016/s2213-2600(20)30116-8]
9. Singh AK, Gupta R, Misra A. Comorbidities in COVID-19: Outcomes in hypertensive cohort and controversies with renin angiotensin system blockers. Diabetes Metab Syndr 2020 Jul;14(4):283-287 [FREE Full text] [doi: 10.1016/j.dsx.2020.03.016] [Medline: 32283499]
10. Jennings G. Coronavirus disease 2019 (COVID-19): angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers and cardiovascular disease. Med J Aust 2020 Jun;212(11):502-503.e1 [FREE Full text] [doi: 10.5694/mja2.50622] [Medline: 32441063]
11. Hippisley-Cox J, Tan PS, Coupland C. Risk of severe COVID-19 disease with ACE inhibitors and angiotensin receptor blockers: cohort study including 8.3 million people. Heart 2020 Nov 05;107(2):170.2-17171. [doi: 10.1136/heartjnl-2020-318312] [Medline: 33153994]
12. Ruocco G, Feola M, Palazzuoli A. Hypertension prevalence in human coronavirus disease: the role of ACE system in infection spread and severity. Int J Infect Dis 2020 Jun;95:373-375 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.058] [Medline: 32335337]
13. Mendes N, Jara C, Mansour E, Araujo E, Velloso L. Asthma and COVID-19 - A systematic review. Authorea 2020:1-3 [FREE Full text] [doi: 10.22541/au.159118771.11841404]
14. Carli G, Cecchi L, Stebbing J, Parronchi P, Farsi A. Is asthma protective against COVID-19? Allergy 2021 Mar 17;76(3):866-868 [FREE Full text] [doi: 10.1111/all.14426] [Medline: 32479648]
15. Weule G. I have asthma. Am I more at risk of having a severe coronavirus infection? ABC News. 2020 Apr 6. URL: https://www.abc.net.au/news/health/2020-04-07/asthma-and-risk-of-severe-coronavirus-infection/12117680 [accessed 2021-02-26]
16. Hanrahan C. Coronavirus poses higher risk for those with diabetes and heart disease — not just older Australians. ABC News. 2020 Apr 5. URL: https://www.abc.net.au/news/2020-04-06/coronavirus-risk-diabetes-heart-disease-covid19/12113264 [accessed 2021-02-26]
17. McCaffery K, Dodd R, Cvejic E, Ayrek J, Batcup C, Isautier JM, et al. Health literacy and disparities in COVID-19-related knowledge, attitudes, beliefs and behaviours in Australia. Public Health Res Pract 2020 Dec 09;30(4) [FREE Full text] [doi: 10.17061/phrp30342012] [Medline: 33294907]
18. Wolf MS, Serper M, Opsasnick L, O'Conor RM, Curtis LM, Benavente JY, et al. Awareness, Attitudes, and Actions Related to COVID-19 Among Adults With Chronic Conditions at the Onset of the U.S. Outbreak. Annals of Internal Medicine 2020 Jul 21;173(2):100-109. [doi: 10.7326/m20-1239]
19. Wallace LS, Rogers ES, Roskos SE, Holiday DB, Weiss BD. Brief report: screening items to identify patients with limited health literacy skills. J Gen Intern Med 2006 Aug;21(8):874-877 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00532.x] [Medline: 16881950]
20. Wolf MS, Smith SG, Pandit AU, Condon DM, Curtis LM, Griffith J, et al. Development and Validation of the Consumer Health Activation Index. Med Decis Making 2018 Feb 13;38(3):334-343. [doi: 10.1177/0272989x17753392]
21. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep 12;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
22. Stepner M, Garland A. CALIPMATCH: Stata module for caliper matching without replacement. EconPapers. 2017 May 11. URL: https://econpapers.repec.org/RePEc:boc:bocode:s458342 [accessed 2021-02-16]
23. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004 Apr 01;159(7):702-706. [doi: 10.1093/aje/kwh090] [Medline: 15033648]

24. Household Impacts of COVID-19 Survey. Australian Bureau of Statistics. 2020 May 1. URL: https://www.abs.gov.au/statistics/people/people-and-communities/household-impacts-covid-19-survey/14-17-apr-2020 [accessed 2021-02-26]
25. Seale H, Heywood AE, Leask J, Sheel M, Durrheim DN, Bolsewicz K, et al. Examining Australian public perceptions and behaviors towards a future COVID-19 vaccine. BMC Infect Dis 2021 Jan 28;21(1):120 [FREE Full text] [doi: 10.1186/s12879-021-05833-1] [Medline: 33509104]
26. Faasse K, Newby J. Public Perceptions of COVID-19 in Australia: Perceived Risk, Knowledge, Health-Protective Behaviors, and Vaccine Intentions. Front Psychol 2020 Sep 30;11:551004 [FREE Full text] [doi: 10.3389/fpsyg.2020.551004] [Medline: 33117223]
27. Dodd RH, Cvejic E, Bonner C, Pickles K, McCaffery KJ, Ayre J, et al. Willingness to vaccinate against COVID-19 in Australia. The Lancet Infectious Diseases 2021 Mar;21(3):318-319. [doi: 10.1016/s1473-3099(20)30559-4]
28. Peretti-Watel P, Seror V, Cortaredona S, Launay O, Raude J, Verger P, et al. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. The Lancet Infectious Diseases 2020 Jul;20(7):769-770 [FREE Full text] [doi: 10.1016/S1473-3099(20)30426-6] [Medline: 32445713]
29. Rhodes A, Hoq M, Measey M, Danchin M. Intention to vaccinate against COVID-19 in Australia. The Lancet Infectious Diseases 2020 Sep. [doi: 10.1016/s1473-3099(20)30724-6]
30. McInerney M. Facebook action creates public health emergency in Australia. Croakey. 2021 Feb 18. URL: https://www.croakey.org/facebook-action-creates-public-health-emergency-in-australia/ [accessed 2021-02-26]

Abbreviations

aOR: adjusted odds ratio
aPR: adjusted prevalence ratio
CHAI: Consumer Health Activation Index
CVD: cardiovascular disease
MMD: marginal mean difference
STAI: State-Trait Anxiety Inventory

Edited by E Meinert; submitted 14.12.20; peer-reviewed by J Delgado-Ron, D Neupane; comments to author 14.01.21; revised version received 02.03.21; accepted 08.03.21; published 30.03.21.

Please cite as:
Bonner C, Cvejic E, Ayre J, Isautier J, Semsarian C, Nickel B, Batcup C, Pickles K, Dodd R, Cornell S, Copp T, McCaffery KJ
The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study
JMIRx Med 2021;2(1):e25610
URL: https://xmed.jmir.org/2021/1/e25610
doi:10.2196/25610
PMID:34076628

Original Paper

Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis

Nader Alharbi[1], PhD

King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Corresponding Author:
Nader Alharbi, PhD
King Saud bin Abdulaziz University for Health Sciences
King Abdullah International Medical Research Center
Prince Mutib Ibn Abdullah Ibn Abdulaziz Rd, Ar Rimayah
Riyadh
Saudi Arabia
Phone: 966 114299999 ext 95590
Email: alharbina@ksau-hs.edu.sa

Related Articles:

Companion article: https://preprints.jmir.org/preprint/21044

Companion article: https://med.jmirx.org/2021/1/e28679/

Companion article: https://med.jmirx.org/2021/1/e28741/

Companion article: https://med.jmirx.org/2021/1/e28742

Abstract

Background: Infectious disease is one of the main issues that threatens human health worldwide. The 2019 outbreak of the new coronavirus SARS-CoV-2, which causes the disease COVID-19, has become a serious global pandemic. Many attempts have been made to forecast the spread of the disease using various methods, including time series models. Among the attempts to model the pandemic, to the best of our knowledge, no studies have used the singular spectrum analysis (SSA) technique to forecast confirmed cases.

Objective: The primary objective of this paper is to construct a reliable, robust, and interpretable model for describing, decomposing, and forecasting the number of confirmed cases of COVID-19 and predicting the peak of the pandemic in Saudi Arabia.

Methods: A modified singular spectrum analysis (SSA) approach was applied for the analysis of the COVID-19 pandemic in Saudi Arabia. We proposed this approach and developed it in our previous studies regarding the separability and grouping steps in SSA, which play important roles in reconstruction and forecasting. The modified SSA approach mainly enables us to identify the number of interpretable components required for separability, signal extraction, and noise reduction. The approach was examined using different levels of simulated and real data with different structures and signal-to-noise ratios. In this study, we examined the capability of the approach to analyze COVID-19 data. We then used vector SSA to predict new data points and the peak of the pandemic in Saudi Arabia.

Results: In the first stage, the confirmed daily cases on the first 42 days (March 02 to April 12, 2020) were used and analyzed to identify the value of the number of required eigenvalues (r) for separability between noise and signal. After obtaining the value of r, which was 2, and extracting the signals, vector SSA was used to predict and determine the pandemic peak. In the second stage, we updated the data and included 81 daily case values. We used the same window length and number of eigenvalues for reconstruction and forecasting of the points 90 days ahead. The results of both forecasting scenarios indicated that the peak would

occur around the end of May or June 2020 and that the crisis would end between the end of June and the middle of August 2020, with a total number of infected people of approximately 330,000.

Conclusions: Our results confirm the impressive performance of modified SSA in analyzing COVID-19 data and selecting the value of *r* for identifying the signal subspace from a noisy time series and then making a reliable prediction of daily confirmed cases using the vector SSA method.

(JMIRx Med 2021;2(1):e21044) doi:10.2196/21044

KEYWORDS

COVID-19; prediction; singular spectrum analysis; separability; eigenvalues; Saudi Arabia

Introduction

One of the main issues that threatens human health worldwide is infectious diseases. Recently, the 2019 outbreak of the new coronavirus, SARS-CoV-2, which causes the disease known as COVID-19, has led to a global pandemic [1,2]. The first case of the virus was recognized and reported on December 31, 2019, in the city of Wuhan, the capital of Hubei Province in China [3]. The virus then spread rapidly worldwide and has affected more than 200 countries [4].

The number of cases and deaths from SARS-CoV-2 globally are considered to be a serious problem [5,6]. As of May 12, 2020, the number of confirmed cases worldwide was more than 4 million, with approximately 200,000 deaths. Although the outbreak appears to have abated in China, the virus and its impact are still spreading globally, and the case numbers are increasing. This is leading to concerns about variations in the affected cases and the mortality rate of the pandemic. Furthermore, there is much concern about the global economic impact of the crisis. It is now understood that the devastating influence of the virus on the economy and world health is without precedent [7].

In addition, several urgent queries related to transmission dynamics, mitigation, and control measures of COVID-19 have been raised, and researchers are attempting to use mathematical modeling to answer these important questions [8]. For example, the containment of transmission, plans such as quarantine, social distancing, and contact tracing of infected or suspected carriers, and lockdowns in regions or countries to address the disease have been included in the results of model predictions [9,10].

There are several standard epidemiological models for modelling epidemics, such as the susceptible, infectious, recovered (SIR) model [11-13]. Many studies have been conducted to model the pandemic using various methods, such as deep learning-based models [14], a simple iteration method [15], generalized additive models [16], which were used to estimate the three parameters of time-dependent transmission, time-dependent recovery, and time-dependent death rates from the outbreak; also, a hybrid model including 2D curvelet transformation, the chaotic salp swarm algorithm, and a deep learning technique was used to identify people infected with SARS-CoV-2 from x-ray images [17].

The primary objective of this study is the construction of a reliable, robust, and interpretable model for describing, decomposing, and forecasting the number of confirmed COVID-19 cases and predicting the peak of the pandemic in Saudi Arabia. The rate of mortality in Saudi Arabia is low, less than 1% at the time of writing this paper (May 12, 2020). Therefore, we were only interested in new daily cases of people affected by SARS-CoV-2 in an attempt to detect its peak. The number of cumulative cases was more than 40,000 as of May 12, 2020.

Because our aim was to analyze the daily data series of COVID-19, we sought to use a promising, reliable, and capable method for analyzing time series. A number of methods can be used to perform such an analysis; however, several of these methods are parametric and thus have requirements such as linearity or nonlinearity of a particular form.

An alternative method is to use nonparametric approaches that are neutral with respect to problematic areas of specification, such as linearity, stationarity, and normality [18]. These approaches can represent a reliable and superior means of decomposing time series data. Singular spectrum analysis (SSA) is a relatively new nonparametric technique that has been proved to be effective in several time series applications in different disciplines, such as genetics and biology [19,20], medicine [21,22], engineering [23,24], and economics and finance [25,26]. For the history of SSA, see [27,28], and for more details on the theory of SSA and its applications, refer to [29,30]. A comprehensive review of the SSA method and descriptions of its extensions and modifications can be found in [31].

The SSA technique is considered to be a useful tool that can be applied to solve many problems, such as smoothing; finding trends in different resolutions; simultaneous extraction of cycles with small and large periods; extraction of seasonality components; extraction of periodicities with varying amplitudes; and simultaneous extraction of complex trends and periodicities [30]. It should be noted that SSA is not linked with generalized autoregressive conditional heteroskedasticity, advanced autoregressive integrated moving average, wavelets, or other methods of this type. However, it has close links with certain methods of multivariate statistics and with signal methods such as projection pursuit and principal component analysis [30,32,33].

Although signals can be affected by internal or external noise, which often has unknown characteristics, they can be identified if the signal and noise subspaces are accurately separated. It is known that removing noise from any signal is necessary for analyzing any time series and is helpful in properly decomposing signals [34].

The main idea of SSA is to analyze the main series into different components, then reconstruct the noise-free series for further analysis. This process depends upon two main choices: the window length L and the number of required eigenvalues, denoted by r, for reconstruction. Therefore, appropriate selection of L and r leads to perfect analysis and separability between the time series components. It was discussed in [35] that for a series of length N, selecting $L=N/4$ is common practice. It should also be mentioned that L needs to be sufficiently large but no larger than half of the series [29]. In [36], it was shown that for a series of length N and the optimal selection of the number of eigenvalues r for reconstructing the signal, the appropriate value of the window length is $median\{1, \ldots, N\}$. Although various attempts have been made, no universal rule has been established for obtaining optimal selections of L and r.

We proposed an approach in [37-39] for the selection of the value of r for noise reduction, filtering, and signal extraction in SSA. This approach has also been applied to the distinction of noise from chaos in time series analysis [40] and for the correction of noise in gene expression data [41]. In [39], we developed the approach and introduced new criteria to the discrimination between epileptic seizure and normal electroencephalogram (EEG) signals, the filtering of the EEG signal segments, and elimination of the noise included in the signal. The approach is mainly used to identify the required number of eigenvalues or singular values corresponding to the signal component, which depends on the distribution of the eigenvalues of a scaled Hankel matrix. The correlation between eigenvalues, the coefficients of skewness, the kurtosis, and the variation of the distribution of the eigenvalues were proposed and proved to be new criteria for the separability between the signal and noise components, as they can split the eigenvalues into two groups [38]. Different simulated and real signals were used to consider different signal-to-noise (SNR) ratios in [38,39] and were evaluated to show the ability of the approach in the selection of r.

The remainder of this paper is structured as follows. The Methods section gives a short description of the modified SSA approach and its algorithm. In the Results section, we show that this approach can be used to decompose synthetic data into two main distinct subspaces, and we then discuss the implementation of the approach in decomposing and reconstructing series of COVID-19 daily cases. This section also presents the forecasting of the COVID-19 pandemic in Saudi Arabia using vector singular spectrum analysis (VSSA) of the signal extracted by modified SSA. The Discussion section draws the conclusion of the paper and suggests ideas for future work.

Methods

The Modified SSA Method: Review

This section presents a short description of the modified SSA used in this manuscript (for more details, refer to [38]). A time series was decomposed by the technique into a sum of components, allowing for identification of each as either a main or noise component. The goal was to consider the signal as a whole so that we could identify the appropriate value of r related to the whole signal component. In other words, we were not interested in each signal component; thus, the selection of L rational to the periodicity of the signal components was less important [30]. Therefore, the modified SSA method focused on the selection of r to identify the signal subspace.

Consider a one-dimensional series $Y_N = (y_1, \ldots, y_N)$ of length N. Transferring this series into a multidimensional series $X_1, \ldots, X_K$, where $X_i = (y_1, \ldots, y_{i+L-1})^T \in \boldsymbol{R}^L$ provides  , where L is an integer ($2 \le L \le N/2$) and $K = N - L + 1$.

A matrix $\boldsymbol{X}$ is a Hankel matrix, in which all the elements along the diagonal $I + j = const$ are equal. Set $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{X}^T$, denote by λ_i ($i = 1, \ldots, L$) the eigenvalues of $\boldsymbol{B}$ taken in decreasing order of magnitude ($\lambda_1 \ge \lambda_L \ge 0$), and denote by $U_1, \ldots, U_L$ the orthonormal system of the eigenvectors of matrix $\boldsymbol{B}$ corresponding to these eigenvalues. The singular value decomposition (SVD) of matrix $\boldsymbol{X}$ can be written as follows:

$$\boldsymbol{X} = \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_L \qquad (1)$$

where  . The elementary matrices $\boldsymbol{X}_i$ having rank 1, $\boldsymbol{U}_i$, and $\boldsymbol{V}_i$ are the left and right eigenvectors of matrix $\boldsymbol{X}$. Note that the collection  is called the i^{th} eigentriple of the SVD. Note also that  and  , where $\| \ \|_F$ denotes the Frobenius norm.

Fundamental to the question of eigenvalue behavior, λ_i, is that if the series size increases, there is a corresponding increase in the eigenvalues. This problem can be overcome if $\boldsymbol{B}$ is divided by its trace, $\boldsymbol{A} = \boldsymbol{B}/tr(\boldsymbol{B})$, which provides several important properties [37]. Let $\zeta_1, \ldots, \zeta_L$ denote the matrix $\boldsymbol{B}$ eigenvalues in decreasing order of magnitude ($1 \ge \zeta_1 \ge \cdots \zeta_L \ge 0$). The simulation is performed to obtain the distribution of ζ_1 and to understand the behavior of each eigenvalue. This helps identify the value of r. Here, the goal was to establish the distribution and related forms of ζ_1 that would be used to select the appropriate value of r for removing noise from the COVID-19 series.

It was proved in our previous work [38] that the largest eigenvalue has a positive skewed distribution for a white noise process. Therefore, if $skew(\zeta_c)$ ($c \in \{1, \ldots, L\}$)is the maximum, and the pattern for $skew(\zeta_c)$ to $skew(\zeta_L)$ has the same pattern, the same as that which emerged for the white noise, then the first $r = c$ 1 eigenvalues correspond to the signal and the remaining eigenvalues correspond to the noise. A similar procedure can be performed using the coefficients of kurtosis and the variation of ζ_i. Furthermore, if $\rho_S(\zeta_{c-1}, \zeta_c)$ is the minimum, and the pattern for the set  is similar to what was observed for the white noise, then we select the first r eigenvalues for the signal and the remainder for the noise component (for more information, see [38]).

In this research, we used the third and fourth central measure moments of the distribution, which are the skewness (*skew*) and kurtosis (*kurt*). Skewness is a measure of asymmetry of the data distribution, while kurtosis describes the distribution of observed data in terms of shape or peak. We used these measures as

criteria for choosing the value of r, which can be calculated for a simulation m as follows:

Moreover, the coefficient of variation (CV), which is defined as the ratio of the standard deviation $\sigma(\zeta_i)$ and  , can be calculated mathematically from the following formula:

In addition, the Spearman correlation ρ_S between the eigenvalues ζ_i and ζ_j ($i, j = 1, \ldots, L$) was calculated to enhance the results obtained by those measures:

where $d_n = x_n - y_n$ ($n = 1, \ldots, m$) is the difference between x_n and y_n, which are the ranks of ζ_i and ζ_j, respectively, and $\zeta_{i,n}$ is the n-th observation for the i-th eigenvalue (ζ_j),  .

These measures of difference between the eigenvalues related to the signal and noise components can specify the cutoff point of separability, namely, the number of leading SVD components that are separated from the residual. Therefore, the final cutoff point of separability between the signal and noise components obtained by the suggested measures corresponds to the rank estimation.

The eigenvalues can be split into two groups by using the above criteria; the first group corresponds to the signal, and the second corresponds to the noise component. Furthermore, the Spearman correlation ρ_S between ζ_i and ζ_j was calculated to support the outcomes obtained by those measures. The absolute value of the correlation coefficient was considered; 1 shows that ζ_i and ζ_j have a perfect positive correlation, while 0 indicates there is no correlation between them. The matrix of the absolute values of the Spearman correlation gives a full analysis of the trajectory matrix, and in this analysis, each eigenvalue corresponds to an elementary matrix of the SVD. Note that if the absolute value of ρ_S is close to 0, the corresponding components are almost orthogonal; however, if it is close to 1, the two components are far from being orthogonal, and thus it is difficult to separate them. Therefore, if ρ_S=0 between two reconstructed components, these two reconstructed series are separable. The results of ρ_S between the eigenvalues for the white noise are quite large (see [38]), which aids the discrimination of the noise part.

Once r is identified, the matrices $\boldsymbol{X}_i$ can be split into two groups. Therefore, Equation 1 can be written as

$$\boldsymbol{X} = \boldsymbol{S} + \boldsymbol{E} \qquad (6)$$

where  is the signal matrix and  is the noise matrix. We then use diagonal averaging to transform matrix $\boldsymbol{S}$ into a new series of size N (see [29]).

The Algorithm

The algorithm consisted of two main stages. The steps in the first stage used the coefficients of skewness, kurtosis, variation, and correlation to help obtain the optimal value of r for the separability between signal and noise, as these coefficients split the eigenvalues into two groups. The steps in the second stage were used to reconstruct the free noise series.

The steps in Stage 1 are outlined below:

1. Map a one-dimensional time series $Y_N = y_1, \ldots, y_N$ into s multidimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1}) \in \boldsymbol{R}^L$, where the window length L is an integer; $2 \leq L \leq N/2$, and $K = N - L + 1$. This step gives us the Hankel matrix  .
2. Compute the matrix $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}^T/tr(\boldsymbol{X}\boldsymbol{X}^T)$.
3. Decompose matrix $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{P}\Gamma\boldsymbol{P}^T$, where $\Gamma = diag(\zeta_i, \ldots, \zeta_L)$ is the diagonal matrix of the eigenvalues of $\boldsymbol{A}$ that has the order ($1 \geq \zeta_i, \ldots, \zeta_L \geq 0$) and $P = P_1, \ldots, P_L$ is an orthogonal matrix whose columns are the corresponding eigenvectors.
4. Simulate the original series m times and calculate the eigenvalues for each series. We simulate y_i from a uniform distribution with boundaries $y_i - a$ and $y_i - b$, where $a = |y_{i-1} - y_i|$ and $b = |y_i - y_{i+1}|$.
5. Compute the skewness coefficient for each eigenvalue, $skew(\zeta_i)$. If $skew(\zeta_c)$ is the maximum, and the pattern for $skew(\zeta_c)$ to $skew(\zeta_L)$ has a similar pattern to that of the white noise, select $r = c - 1$.
6. Compute the coefficient of kurtosis for each eigenvalue, $kurt(\zeta_i)$. If $skew(\zeta_c)$ is the maximum, select $r = c - 1$.
7. Compute the coefficient of variation, CV=ζ_i. The result of the CV splits the eigenvalues into two groups; the eigenvalues from ζ_i to ζ_{c-1} correspond to the signal, and the remaining eigenvalues, which have an almost U shape, correspond to the noise.
8. Compute the absolute values of the correlation matrix between the eigenvalues and represent them in a 20-grade grey scale from white to black corresponding to the values of the correlations from 0 to 1. This matrix also splits the eigenvalues into two groups; the eigenvalues from ζ_i to ζ_r correspond to the signal, and the remaining eigenvalues correspond to the noise.

The steps in Stage 2 are outlined below:

1. Calculate the approximated signal matrix  , that is,  , where r is obtained from the first stage, and  , where U_i and V_i represent the left and right eigenvectors of the trajectory matrix, respectively.
2. Averaging over the diagonals of the matrix  gives a one-dimensional series, which is the approximate signal  .

The capabilities of modified SSA using different types of synthetic data, including series generated from chaotic map

systems with different SNR ratios, are presented in [38]. This study confirms that the approach works promisingly for any series that is mixed with a low or high noise level.

Each eigenvalue or singular value contributes to the trajectory matrix decomposition. We can consider the ratio to be the characteristic of matrix H_i to Equation 1. Therefore, [illegible] is considered to be characteristic of the optimal approximation of $\boldsymbol{H}$ by matrices of rank r.

Results

Separability in Synthetic Data

It should be noted that using the standard criteria in basic SSA, the weighted correlation (*w*-correlation) for separability and grouping (for more information, see [29]), does not always provide good separability and correct selection of *r*, especially for real data.

It was shown in [38] that the results based on *skew*, *kurt*, *CV*, and ρ_S are more accurate than those obtained by the *w*-correlations for small window lengths, particularly for data in which a linear trend is included in the series.

We therefore used modified SSA—in particular, some of the proven criteria on the distribution of ζ_i, as given in the previous sections—to identify *r*. The results were plausible and reliable.

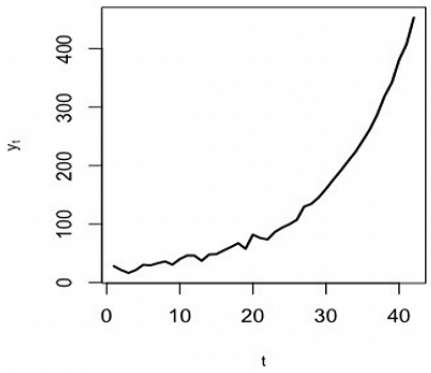
Below, we provide a synthetic example to show the capability of the approach before applying it to the COVID-19 data; for more examples considering different types of series and evaluations with different criteria, refer to [38].

In the following example, a white noise process [illegible] was added to an exponential trend series:

$$Y_t = \alpha_1 + \alpha_2 \exp(\alpha_2 t) + [illegible] \quad \textbf{(7)}$$

where $t=(1, \ldots, N)$, $N=42$, $\alpha_1=10$, $\alpha_2=0.09$, and [illegible] is a Gaussian white noise process with variance 1 (see Figure 1). It is obvious that the number of eigenvalues required to reconstruct the signal for this series is 2, as we have added a constant to the exponential curve, which corresponds to the rank estimation (see [29]).

Figure 1. Realization of the simulated exponential trend series.



Based on observations of the *w*-correlations and the logarithm of the eigenvalues, one may use only the first component to extract the signal (see Figure 2). However, using the suggested measures and criteria gives the correct value of *r*. Figure 3 shows the kurtosis coefficient of ζ_i ($i=1, \ldots, L$). The maximum value of the kurtosis coefficient is considered as one of the rules and indicators used for the start of the noise. It is clear that the maximum kurtosis coefficient of ζ_i is obtained for $\zeta_{c=3}$. Therefore, the number of eigenvalues required to extract the signal is $r = c - 1 = 2$. Similar results were obtained using the values of *skew* and *CV* (see Figure 4).

Figure 2. Left: w-correlation matrix for the seven reconstructed components of the simulated series. Right: logarithms of the seven eigenvalues of the simulated series. w-correlation: weighted correlation.

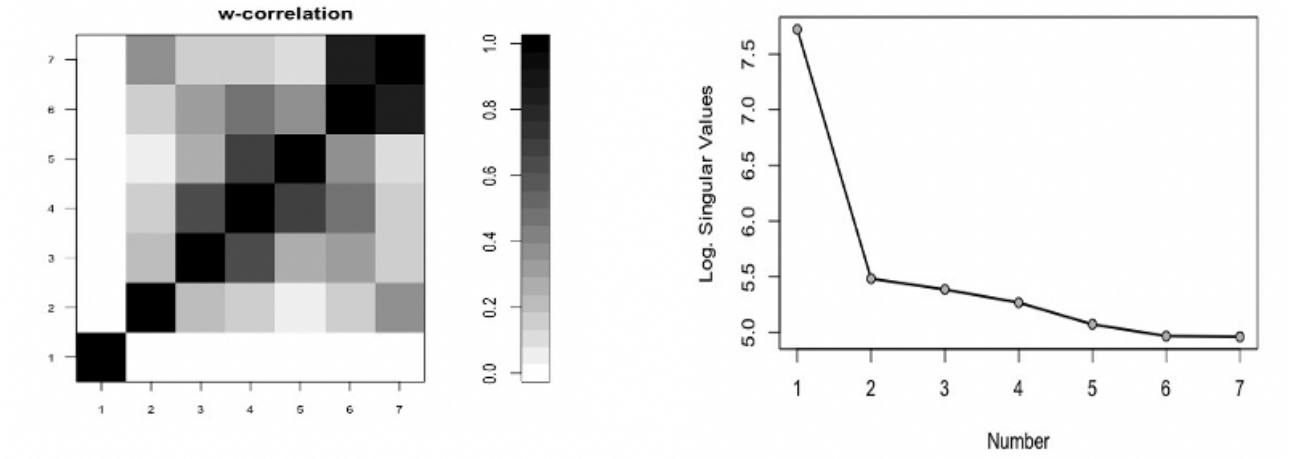


Figure 3. Kurt of ζ_i for the simulated series. Kurt: kurtosis.

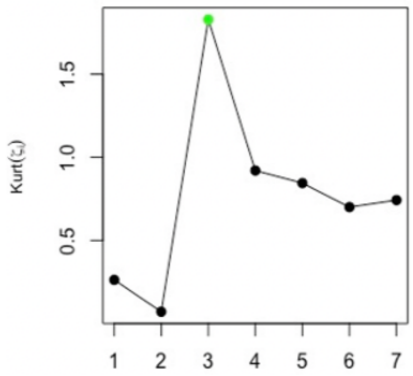
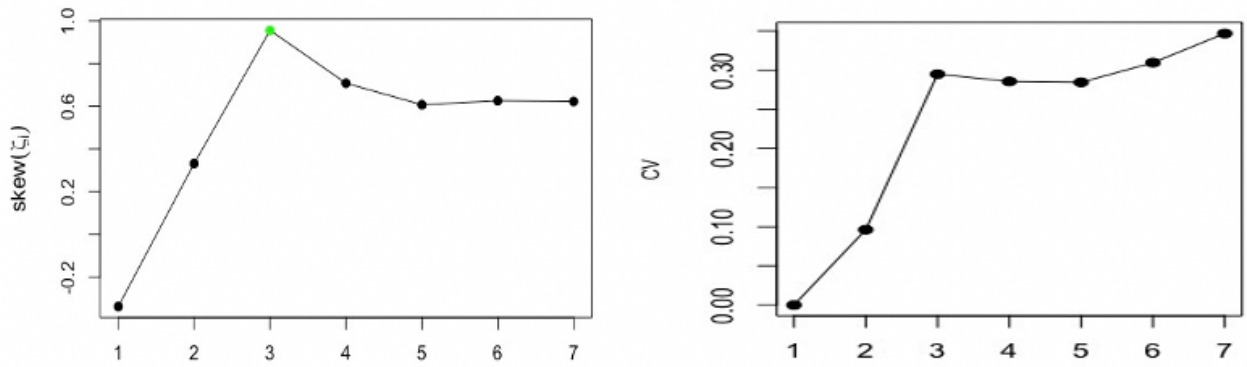
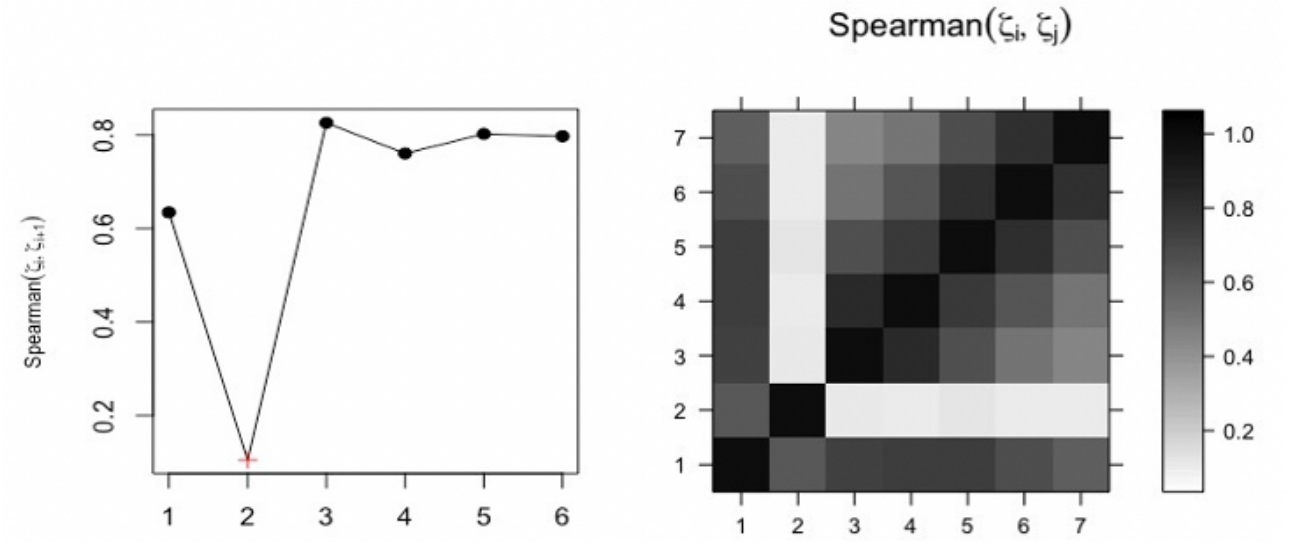


Figure 4. Left: skew of ζ_i for the simulated series. Right: CVs of ζ_i for the simulated series. CV: coefficient of variation; skew: skewness.



In addition, the Spearman correlation coefficient between ζ_i and ζ_{i+1} was calculated; Figure 5 (left) shows the correlation between ζ_i and ζ_{i+1}. For the correlation coefficient, the minimum value of ρ_S between ζ_i and ζ_{i+1} was used as another indicator for the cutoff point. The results were similar to those that emerged using other criteria and confirmed that the approach works properly. Different criteria, such as root mean square error and mean absolute error, were used in [38] to evaluate the approach, and the results confirmed that the modified approach is a promising one.

Figure 5. Left: Spearman correlation of (ζ_i,$\zeta_{(i+1)}$). Right: matrix of Spearman correlation between (ζ_i,ζ_j).

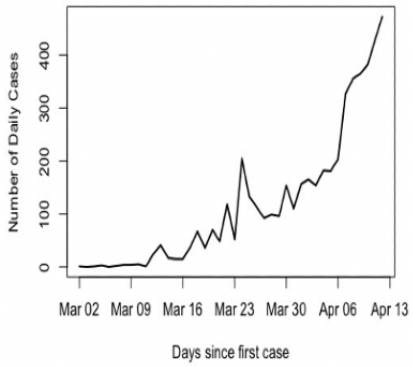


The correlation matrix also enables us to distinguish and separate the different components from each other. Therefore, the correlation matrix of ζ_i identifies the separability between the components. If the absolute value of the correlation coefficient between ζ_i and ζ_j is small, then the corresponding components are almost orthogonal; however, if the value is large, then the corresponding series are far from being orthogonal, and thus they are not neatly separable. It is clear that the signal can be separated from the noise, as the top right-hand pattern from the correlation matrix is related to the white noise process (see Figure 5, right).

COVID-19 Data Analysis

The daily numbers of confirmed cases of COVID-19 in Saudi Arabia [42] were used in this research. First, we used data from the first 42 days, from March 2 to April 12, 2020. The aim was to analyze the data, make predictions from April 13, 2020, and detect the peak. The number of daily cases series is shown in Figure 6. Second, we updated our data on May 20, 2020, to include values from April 13 to May 12, 2020; thus, the total became 81 values. This did not affect the required number of eigenvalues for the reconstruction stage, as will be discussed in the following section.

Figure 6. Time series of daily confirmed COVID-19 cases in Saudi Arabia (March 2 to April 12, 2020).



Separability and Selection of the Components

Starting with the first set of COVID-19 data, as mentioned earlier, because our aim was to extract the signal as a whole, we could choose any value for L, with the goal to find the best choice of r. Furthermore, in our previous research [38], we showed that it is possible to use a small window length when analyzing exponential series, like the series of COVID-19 cases. The selection of L=7 provided the best and most reasonable results with the required r that would be obtained by the proposed approach.

The results based on these measures in extracting the signal for forecasting gave a curve with a likely peak. However, the predictions using various other choices for L and r did not indicate any end or peak for the pandemic and in fact showed exponential increases; such increases are impossible, as the pandemic will not continue forever. This finding also supports the obtained results. Therefore, the next important task was the selection of the number of eigenvalues r required for the reconstruction and building of the model for forecasting.

Figure 7 illustrates the coefficients of skewness and kurtosis for each eigenvalue and the results of the matrix correlations and the correlations between ζ_i and ζ_{i+1} for L=7. As shown by the results, for the COVID-19 daily series, the maximum values of *skew* and *kurt* are observed for $\zeta_{c=3}$, and the minimum value of ρ_S is obtained between $\zeta_{c-1=3}$ and $\zeta_{c=3}$. In addition, the matrix of the Spearman correlation for ζ_i and ζ_j splits the eigenvalues

or the components into two groups, which indicates that the value of *r* is 2.

Figure 7. Coefficients of skewness (top left) and kurtosis (top right) for each eigenvalue and the correlations between ζ_i and ζ_{i+1} (bottom left) and the results of the matrix correlations (bottom right) for L=7.

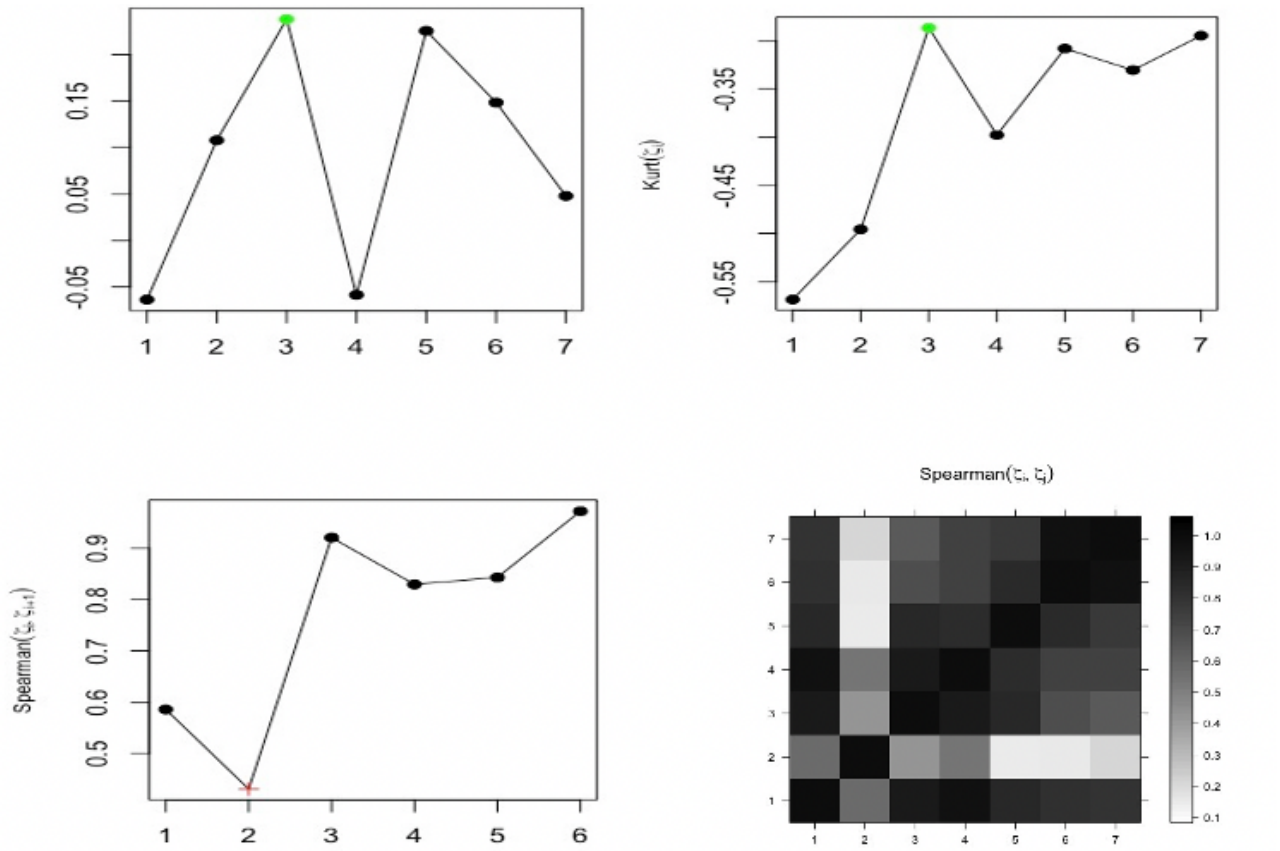
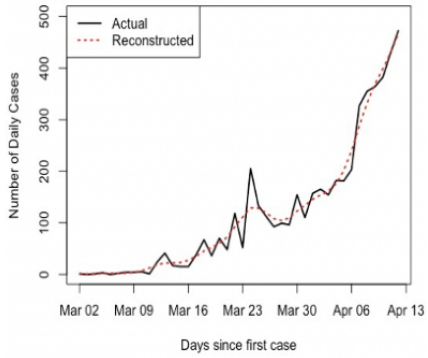


Figure 8 shows the results of the reconstructed series obtained by using *L*=7 and eigentriples *r*=2. The red and black lines correspond to the reconstructed series and the original series, respectively. It appears that the reconstructed series that was obtained is good. However, it will be shown later that the reconstructed series using the whole data set is better than this fitted series.

Figure 8. Plot of the first time series of daily COVID-19 cases in Saudi Arabia and the fitted curve.



Prediction of Daily Cases of COVID-19 Using VSSA

After obtaining the reconstructed series, the next aim was to predict the data for daily new cases from April 13 to August 2020. There are two main forecasting methods in SSA: VSSA (VSSA) and recurrent singular spectrum analysis (RSSA). The VSSA forecasting algorithm is the most widely used in SSA [29]. Generally, this method is more robust than RSSA, especially when a series contains outliers or when facing large shocks in the series [43]. Therefore, we focused on the use of the VSSA algorithm for forecasting in this research, as recommended in [18].

Vector Forecasting Algorithm

To perform SSA forecasting, the basic requirement is that the series satisfies a linear recurrent formula (LRF). The series $Y_N = [y_1, \ldots, y_N]$ satisfies an LRF of order L 1 if

$$Y_t = a_1 y_{t-1} + a_2 y_2 + \cdots + a_{L-1} y_{t-L+1},\ t = L + 1, \ldots, N \quad \textbf{(8)}$$

The coefficient vector $A = a_1, \ldots, a_{L-1}$ is defined as follows:

where  ,  is the vector of the first $L-1$ components of the eigenvector U_j, and π_j is the last component of U_j ($j = 1, \ldots, r$).

Consider the following matrix:

$$\Pi = \boldsymbol{U}^{\nabla}\boldsymbol{U}^{\nabla T} + (1 - v^2)\boldsymbol{A}\boldsymbol{A}^T \quad \textbf{(10)}$$

Let us now define the linear operator:

where  = $span\{U_1, \ldots, U_r\}$ and

where Y_Δ is the vector of the last $L-1$elements of Y_N. The vector Z_j is defined as follows:

where  are the reconstructed columns of the trajectory matrix of the i-th series after grouping and leaving out noise components. Now, by constructing matrix $\boldsymbol{Z} = [Z_1, \ldots, Z_{K+h+L-1}]$ and performing diagonal averaging, a new series  is obtained, where  from the h terms of the VSSA forecast.

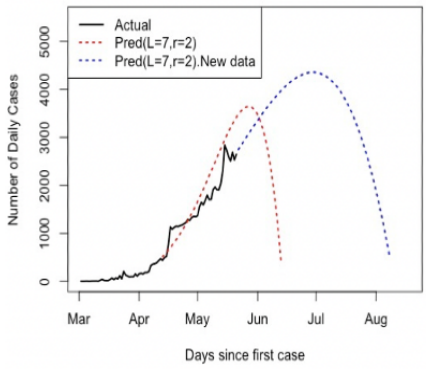
As discussed above, the best values for reconstruction were L=7 and r=2. The values of L=6 and r=3 were the second-best choices based on the criteria presented earlier. For forecasting, the results of these two choices were compared by using the complement statistical test introduced in [44], which is proposed for distinguishing between the predictive accuracy of two sets of forecasts. It is a nonparametric test founded upon the principles of the Kolmogorov-Smirnov test and known as the KS predictive accuracy (KSPA) test. The test is useful for serving two different purposes. First, 2-sided KSPA is used to determine if there is a statistically significant difference between the distribution of forecast errors. Second, the 1-sided KSPA test exploits the principles of stochastic dominance to determine whether the forecasts with lower error also produce a stochastically smaller error than forecasts from a competing model, and it then allows for differentiation between the predictive accuracy of the forecasts [45].

The 2-sided KSPA test indicated that there was no statistically significant difference between the distribution of forecast errors at a 95% confidence level (P=.56). Moreover, there was insufficient evidence based on the one-sided KSPA test at the 5% significance level to conclude that the stochastic errors are different (P=.76). Therefore, the results confirm that there is no statistically significant difference between the two forecasts.

Consequently, we also concentrated only on the best values obtained, L=7 and r=2, for forecasting. Similar procedures were followed for the new data updated on May 20, 2020. The same values of L and r were used to analyze the new data and also for predicting confirmed cases 3 months ahead. Figure 9 shows the updated data and the reconstructed series by the first two eigentriples. It is obvious that the reconstructed series was obtained precisely. Figure 10 shows the two curve predictions and the overall actual data; the red curve is the prediction using the first set of data, and the blue curve is the prediction using the updated data set. It is clear that there is no great difference between the two curves, as the peak appears around the end of May in the red curve and toward the end of June in the blue curve, which was obtained using the updated data. In addition, the end of the pandemic is predicted to occur between July and the middle of August, with the total number of infected people at approximately 330,000.

Figure 9. Plot of the entire time series of daily COVID-19 cases in Saudi Arabia and the fitted curve.

Figure 10. Comparison of the two forecasting scenarios with actual observations. Pred: predicted.



Discussion

A modified SSA approach was used in this research for the decomposition and forecasting of COVID-19 data in Saudi Arabia. The approach was examined in our previous research and was applied here to the analysis of COVID-19 data.

In the first stage, the first 42 values of confirmed daily cases (March 2 to April 12, 2020) were used and analyzed to identify the value of *r* for separability between the noise and signal. After obtaining the value of *r*, which was 2, and extracting the signals, VSSA was used for the prediction and determination of the pandemic peak. In the second stage, we updated the data and included 81 daily values. We used the same window length and number of eigenvalues for the reconstruction and forecasting of the points 90 days ahead. The results of both forecasting scenarios indicated that the peak would occur around the end of May or June and the crisis would end between the end of June and the middle of August 2020, with a total number of infected people of approximately 330,000.

All our results confirm the impressive performance of modified SSA in analyzing the COVID-19 data and selecting the value of *r* for identifying the signal subspace from a noisy time series, then making an accurate prediction using the VSSA method. Note that we did not examine all possible window length values in this research, and for forecasting, we only used basic VSSA.

In future research, we will include more data and consider different window lengths *L*, which may provide better forecasting. In addition, chaotic behavior in the COVID-19 data will be examined, as some of our results show strange patterns, as can be found in chaotic systems.

Conflicts of Interest

None declared.

References

1. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2020-05-02]
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020 Feb 22;395(10224):565-574 [FREE Full text] [doi: 10.1016/S0140-6736(20)30251-8] [Medline: 32007145]
3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
4. Zhou G, Chen S, Chen Z. Back to the spring of 2020: facts and hope of COVID-19 outbreak. Front Med 2020 Apr 14;14(2):113-116. [doi: 10.1007/s11684-020-0758-9] [Medline: 32172487]
5. Goodman-Casanova JM, Dura-Perez E, Guzman-Parra J, Cuesta-Vargas A, Mayoral-Cleries F. Telehealth home support during COVID-19 confinement for community-dwelling older adults with mild cognitive impairment or mild dementia: survey study. J Med Internet Res 2020 May 22;22(5):e19434 [FREE Full text] [doi: 10.2196/19434] [Medline: 32401215]
6. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) model. JMIR Public Health Surveill 2020 May 13;6(2):e19115 [FREE Full text] [doi: 10.2196/19115] [Medline: 32391801]
7. COVID-19 coronavirus pandemic. Worldometer. URL: https://worldometers.info/coronavirus/ [accessed 2020-05-12]
8. Chen T, Rui J, Wang Q, Zhao Z, Cui J, Yin L. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infect Dis Poverty 2020 Feb 28;9(1):24 [FREE Full text] [doi: 10.1186/s40249-020-00640-3] [Medline: 32111262]
9. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
10. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. The effect of control strategies to reduce social mixing on outcomes of the

COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020 May;5(5):e261-e270 [FREE Full text] [doi: 10.1016/S2468-2667(20)30073-6] [Medline: 32220655]
11. Daley D, Gani J. Epidemic Modelling: An Introduction. Cambridge, UK: Cambridge University Press; 2001.
12. Hethcote H. The mathematics of infectious diseases. SIAM Rev 2000 Jan;42(4):599-653. [doi: 10.1137/s0036144500371907]
13. Ferguson N, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London. 2020 Mar 16. URL: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/ [accessed 2020-05-16]
14. Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. Chaos Solitons Fractals 2020 Oct;139:110017 [FREE Full text] [doi: 10.1016/j.chaos.2020.110017] [Medline: 32572310]
15. Perc M, Gorišek Miksić N, Slavinec M, Stožer A. Forecasting COVID-19. Front Phys 2020 Apr 8;8:127. [doi: 10.3389/fphy.2020.00127]
16. Zareie B, Roshani A, Mansournia MA, Rasouli MA, Moradi G. A model for COVID-19 prediction in Iran based on China parameters. Arch Iran Med 2020 Apr 01;23(4):244-248. [doi: 10.34172/aim.2020.05] [Medline: 32271597]
17. Altan A, Karasu S. Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. Chaos Solitons Fractals 2020 Nov;140:110071 [FREE Full text] [doi: 10.1016/j.chaos.2020.110071] [Medline: 32834627]
18. Hassani H, Mahmoudvand R. Singular Spectrum Analysis: Using R. Berlin, Germany: Springer; 2018.
19. Movahedifar M, Yarmohammadi M, Hassani H. Bicoid signal extraction: another powerful approach. Math Biosci 2018 Sep;303:52-61. [doi: 10.1016/j.mbs.2018.06.002] [Medline: 29981354]
20. Ghodsi Z, Silva ES, Hassani H. Bicoid signal extraction with a selection of parametric and nonparametric signal processing techniques. Genomics Proteomics Bioinformatics 2015 Jun;13(3):183-191 [FREE Full text] [doi: 10.1016/j.gpb.2015.02.006] [Medline: 26197438]
21. Sanei S, Hassani H. Singular Spectrum Analysis of Biomedical Signals. Boca Raton, FL: CRC Press; 2015.
22. Safi SMM, Pooyan M, Motie Nasrabadi A. Improving the performance of the SSVEP-based BCI system using optimized singular spectrum analysis (OSSA). Biomed Signal Process Control 2018 Sep;46:46-58. [doi: 10.1016/j.bspc.2018.06.010]
23. Muruganatham B, Sanjith M, Krishnakumar B, Satya Murty S. Roller element bearing fault diagnosis using singular spectrum analysis. Mech Syst Signal Process 2013 Feb;35(1-2):150-166. [doi: 10.1016/j.ymssp.2012.08.019]
24. Liu K, Law S, Xia Y, Zhu X. Singular spectrum analysis for enhancing the sensitivity in structural damage detection. J Sound Vib 2014 Jan;333(2):392-417. [doi: 10.1016/j.jsv.2013.09.027]
25. Hassani H, Rua A, Silva ES, Thomakos D. Monthly forecasting of GDP with mixed-frequency multivariate singular spectrum analysis. Int J Forecast 2019 Oct;35(4):1263-1272. [doi: 10.1016/j.ijforecast.2019.03.021]
26. de Carvalho M, Rodrigues PC, Rua A. Tracking the US business cycle with a singular spectrum analysis. Econ Lett 2012 Jan;114(1):32-35. [doi: 10.1016/j.econlet.2011.09.007]
27. Broomhead D, King GP. Extracting qualitative dynamics from experimental data. Physica D 1986 Jun;20(2-3):217-236. [doi: 10.1016/0167-2789(86)90031-x]
28. Broomhead D, King G. On the qualitative analysis of experimental dynamical systems. In: Nonlinear Phenomena and Chaos. Boston, MA: Adam Hilger Ltd; 1986:113-114.
29. Golyandina N, Nekrutkin V, Zhigljavsky A. Analysis of time series structure. In: SSA and Related Techniques. Boca Raton, FL: CRC Press; 2001.
30. Golyandina N, Zhigljavsky A. Singular Spectrum Analysis for Time Series. Berlin, Germany: Springer; 2013.
31. Golyandina N. Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing. WIREs Comp Stat 2020 Jan 09;12(4):1-39. [doi: 10.1002/wics.1487]
32. Hassani H, Dionisio A, Ghodsi M. The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. Nonlinear Anal Real World Appl 2010 Feb;11(1):492-502. [doi: 10.1016/j.nonrwa.2009.01.004]
33. Hassani H, Yeganegi MR, Khan A, Silva ES. The effect of data transformation on singular spectrum analysis for forecasting. Signals 2020 May 07;1(1):4-25. [doi: 10.3390/signals1010002]
34. Kalantari M, Hassani H. Automatic grouping in singular spectrum analysis. Forecasting 2019 Oct 30;1(1):189-204. [doi: 10.3390/forecast1010013]
35. Elsner J, Tsonis A. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Berlin, Germany: Springer Science & Business Media; 1996.
36. Hassani H, Mahmoudvand R, Zokaei M, Ghodsi M. On the separability between signal and noise in singular spectrum analysis. Fluct Noise Lett 2012 Jul 11;11(02):1250014. [doi: 10.1142/s0219477512500149]
37. Hassani H, Alharbi N, Ghodsi M. A study on the empirical distribution of the scaled Hankel matrix eigenvalues. J Adv Res 2015 Nov;6(6):925-929 [FREE Full text] [doi: 10.1016/j.jare.2014.08.008] [Medline: 26644930]
38. Alharbi N, Hassani H. A new approach for selecting the number of the eigenvalues in singular spectrum analysis. J Franklin Inst 2016 Jan;353(1):1-16. [doi: 10.1016/j.jfranklin.2015.10.015]

39. Alharbi N. A novel approach for noise removal and distinction of EEG recordings. Biomed Signal Process Control 2018 Jan;39:23-33. [doi: 10.1016/j.bspc.2017.07.011]
40. Hassani H, Alharbi N, Ghodsi M. Distinguishing chaos from noise: A new approach. Int J Energy Stat 2014 Jun;02(02):137-150. [doi: 10.1142/s2335680414500100]
41. Alharbi N, Ghodsi Z, Hassani H. Noise correction in gene expression data: a new approach based on subspace method. Math Meth Appl Sci 2016 Mar 08;39(13):3750-3757. [doi: 10.1002/mma.3823]
42. COVID-19 dashboard. Webpage in Arabic. Saudi Arabia Ministry of Health. URL: https://covid19.moh.gov.sa [accessed 2020-05-12]
43. Hassani H, Mahmoudvand R, Omer HN, Silva ES. A preliminary investigation into the effect of outlier(s) on singular spectrum analysis. Fluct Noise Lett 2014 Oct 20;13(04):1450029. [doi: 10.1142/s0219477514500291]
44. Hassani H, Silva E. A Kolmogorov-Smirnov based test for comparing the predictive accuracy of two sets of forecasts. Econometrics 2015 Aug 04;3(3):590-609. [doi: 10.3390/econometrics3030590]
45. Hassani H, Heravi S, Zhigljavsky A. Forecasting European industrial production with singular spectrum analysis. Int J Forecast 2009 Jan;25(1):103-118. [doi: 10.1016/j.ijforecast.2008.09.007]

Abbreviations

***CV*:** coefficient of variation
EEG: electroencephalogram
KSPA: Kolmogorov-Smirnov predictive accuracy
***kurt*:** kurtosis
RSSA: recurrent singular spectrum analysis
SIR: susceptible, infectious, recovered
***skew*:** skewness
SNR: signal-to-noise ratio
SSA: singular spectrum analysis
SVD: singular value decomposition
VSSA: vector singular spectrum analysis
***w*-correlation:** weighted correlation

Edited by E Meinert; submitted 09.06.20; peer-reviewed by Anonymous, Anonymous; comments to author 06.12.20; revised version received 26.12.20; accepted 14.02.21; published 31.03.21.

Please cite as:
Alharbi N
Forecasting the COVID-19 Pandemic in Saudi Arabia Using a Modified Singular Spectrum Analysis Approach: Model Development and Data Analysis
JMIRx Med 2021;2(1):e21044
URL: https://xmed.jmir.org/2021/1/e21044
doi:10.2196/21044
PMID:34076627

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org