Meta-epidemiology of infectious disease models

Bordeaux 10/2022

John P.A. Ioannidis, MD, DSc
Professor of Medicine, of Epidemiology and Population Health, and (by courtesy) of Biomedical Data Science and of Statistics
Co-Director, Meta-Research Innovation Center at Stanford (METRICS)
Stanford University

Infectious disease models and their meta-epidemiology

- Models have a long tradition of successful (or at least insightful) applications in infectious diseases, e.g. SIR model almost a century
- Can be very useful conceptually, with diverse interesting applications and broadening spectrum
- Mixture of data and speculation/assumptions
- Acquired tremendous prominence during the COVID-19 pandemic
- Crash test for models and for science at large
- Used by both highly specialized and well-trained people and by others who jumped into the fray
- Long-standing issues becoame more manifest under the new expedient and high-visibility circumstances

720,000 scientists published scientific papers on COVID-19 indexed by August 1, 2021

(Ioannidis J. et al, Royal Society Open Science 2021)

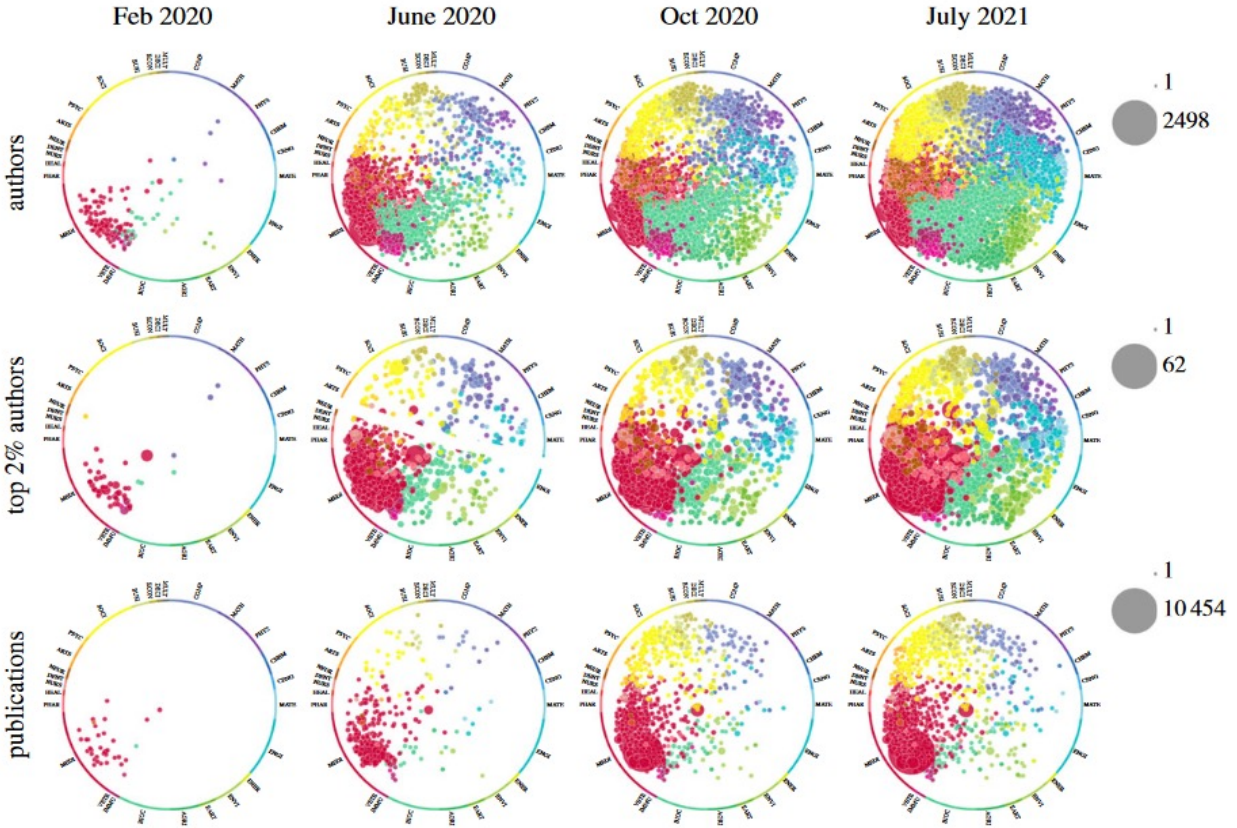


Figure 1. Topics of prominence for COVID-19 authors and publications. The columns represent the progress of the spread at three different measuring points: by end of February 2020, end of June 2020, end of October 2020 and end of July 2021. The first row represents the spread of authors of COVID-19 papers. The authors are assigned to their most dominant topic in their career. The data are filtered to include only topics with greater than or equal to five authors assigned. The second row shows similarly the topics of

PNAS RESEARCH ARTICLE | MEDICAL SCIENCES OPEN ACCESS

Massive covidization of research citations and the citation elite

John P. A. Ioannidis[a,b,c,d,e,1], Eran Bendavid[a], Maia Salholz-Hillel[f], Kevin W. Boyack[g], and Jeroen Baas[h]

Edited by Kenneth Wachter, University of California, Berkeley, CA; received March 7, 2022; accepted May 31, 2022

Massive scientific productivity accompanied the COVID-19 pandemic. We evaluated the citation impact of COVID-19 publications relative to all scientific work published in 2020 to 2021 and assessed the impact on scientist citation profiles. Using Scopus data until August 1, 2021, COVID-19 items accounted for 4% of papers published, 20% of citations received to papers published in 2020 to 2021, and >30% of citations received in 36 of the 174 disciplines of science (up to 79.3% in general and internal medicine). Across science, 98 of the 100 most-cited papers published in 2020 to 2021 were related to COVID-19; 110 scientists received ≥10,000 citations for COVID-19 work, but none received ≥10,000 citations for non–COVID-19 work published in 2020 to 2021. For many scientists, citations to their COVID-19 work already accounted for more than half of their total career citation count. Overall, these data show a strong covidization of research citations across science, with major impact on shaping the citation elite.

Significance

The COVID-19 pandemic saw a massive mobilization of the scientific workforce. We evaluated the citation impact of COVID-19 publications relative to all scientific work published in 2020 to 2021, finding that 20% of citations received to papers published in 2020 to 2021 were to COVID-19–related papers. Across

Massive covidization of science

- 98 of the top-100 most-cited scientific articles published in 2020 were on COVID-19
- Tens of thousands of scientists received more citations to their work in 2020-2021 than they had received in their entire career.
- Among the top-100 ranked scientists across science in 2020-2021, 70 focused on Health Sciences subfields and most (57/70) had risen to such extremely high ranks even though they did not belong to the top-1000 ranked in 2018-2019. 12 of the 70 were editors or journal staff who published profusely in their journals, mostly on COVID-19.
- The massive funding of COVID-19 research will make reversal of science covidization difficult after the end of the pandemic.

Methodological quality of COVID-19 clinical research

Richard G. Jung[1,2,3,13], Pietro Di Santo[1,2,4,5,13], Cole Clifford[6], Graeme Prosperi-Porta[7], Stephanie Skanes[6], Annie Hung[8], Simon Parlow[4], Sarah Visintini[9], F. Daniel Ramirez[1,4,10,11], Trevor Simard[1,2,3,4,12] & Benjamin Hibbert[2,3,4]✉

The COVID-19 pandemic began in early 2020 with major health consequences. While a need to disseminate information to the medical community and general public was paramount, concerns have been raised regarding the scientific rigor in published reports. We performed a systematic review to evaluate the methodological quality of currently available COVID-19 studies compared to historical controls. A total of 9895 titles and abstracts were screened and 686 COVID-19 articles were included in the final analysis. Comparative analysis of COVID-19 to historical articles reveals a shorter time to acceptance (13.0[IQR, 5.0–25.0] days vs. 110.0[IQR, 71.0–156.0] days in COVID-19 and control articles, respectively; $p < 0.0001$). Furthermore, methodological quality scores are lower in COVID-19 articles across all study designs. COVID-19 clinical studies have a shorter time to publication and have lower methodological quality scores than control studies in the same journal. These studies should be revisited with the emergence of stronger evidence.

Scientific quality of COVID-19 and SARS CoV-2 publications in the highest impact medical journals during the early phase of the pandemic: A case control study

Marko Zdravkovic[1✉], Joana Berger-Estilita[2✉], Bogdan Zdravkovic[1], David Berger[3]*

COVID-19-related medical research: a meta-research and critical appraisal

Marc Raynaud[1†], Huanxi Zhang[2†], Kevin Louis[1†], Valentin Goutaudier[1,3†], Jiali Wang[2], Quentin Dubourg[4], Yongcheng Wei[2], Zeynep Demir[1,5], Charlotte Debiais[1], Olivier Aubert[1], Yassine Bouatou[1], Carmen Lefaucheur[6], Patricia Jabre[7], Longshan Liu[2], Changxi Wang[2], Xavier Jouven[1], Peter Reese[1,8], Jean-Philippe Empana[1] and Alexandre Loupy[1*]

Abstract

Background: Since the start of the COVID-19 outbreak, a large number of COVID-19-related papers have been published. However, concerns about the risk of expedited science have been raised. We aimed at reviewing and categorizing COVID-19-related medical research and to critically appraise peer-reviewed original articles.

Methods: The data sources were Pubmed, Cochrane COVID-19 register study, arXiv, medRxiv and bioRxiv, from 01/11/2019 to 01/05/2020. Peer-reviewed and preprints publications related to COVID-19 were included, written in English or Chinese. No limitations were placed on study design. Reviewers screened and categorized studies according to *i)* publication type, *ii)* country of publication, and *iii)* topics covered. Original articles were critically appraised using validated quality assessment tools.

Results: Among the 11,452 publications identified, 10,516 met the inclusion criteria, among which 7468 (71.0%) were peer-reviewed articles. Among these, 4190 publications (56.1%) did not include any data or analytics (comprising expert opinion pieces). Overall, the most represented topics were infectious disease ($n = 2326$, 22.1%), epidemiology ($n = 1802$, 17.1%), and global health ($n = 1602$, 15.2%). The top five publishing countries were China (25.8%), United States (22.3%), United Kingdom (8.8%), Italy (8.1%) and India (3.4%). The dynamic of publication showed that the exponential growth of COVID-19 peer-reviewed articles was mainly driven by publications without original data (mean 261.5 articles ± 51.1 per week) as compared with original articles (mean of 69.3 ± 22.3 articles per week). Original articles including patient data accounted for 713 (9.5%) of peer-reviewed studies. A total of 576 original articles (80.8%) showed intermediate to high risk of bias. Last, except for simulation studies that mainly used large-scale open data, the median number of patients enrolled was of 102 (IQR = 37–337).

Conclusions: Since the beginning of the COVID-19 pandemic, the majority of research is composed by publications without original data. Peer-reviewed original articles with data showed a high risk of bias and included a limited number of patients. Together, these findings underscore the urgent need to strike a balance between the velocity and quality of research, and to cautiously consider medical information and clinical applicability in a pressing, pandemic context.

(Continued on next page)

Why might models fail?

European Journal of Epidemiology (2020) 35:733–742
https://doi.org/10.1007/s10654-020-00669-6

COVID-19

Check for updates

A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York state

Vincent Chin[1,2] · Noelle I. Samia[3] · Roman Marchant[1,2] · Ori Rosen[4] · John P. A. Ioannidis[5,6,7,8,9,10] · Martin A. Tanner[3] · Sally Cripps[1,2]

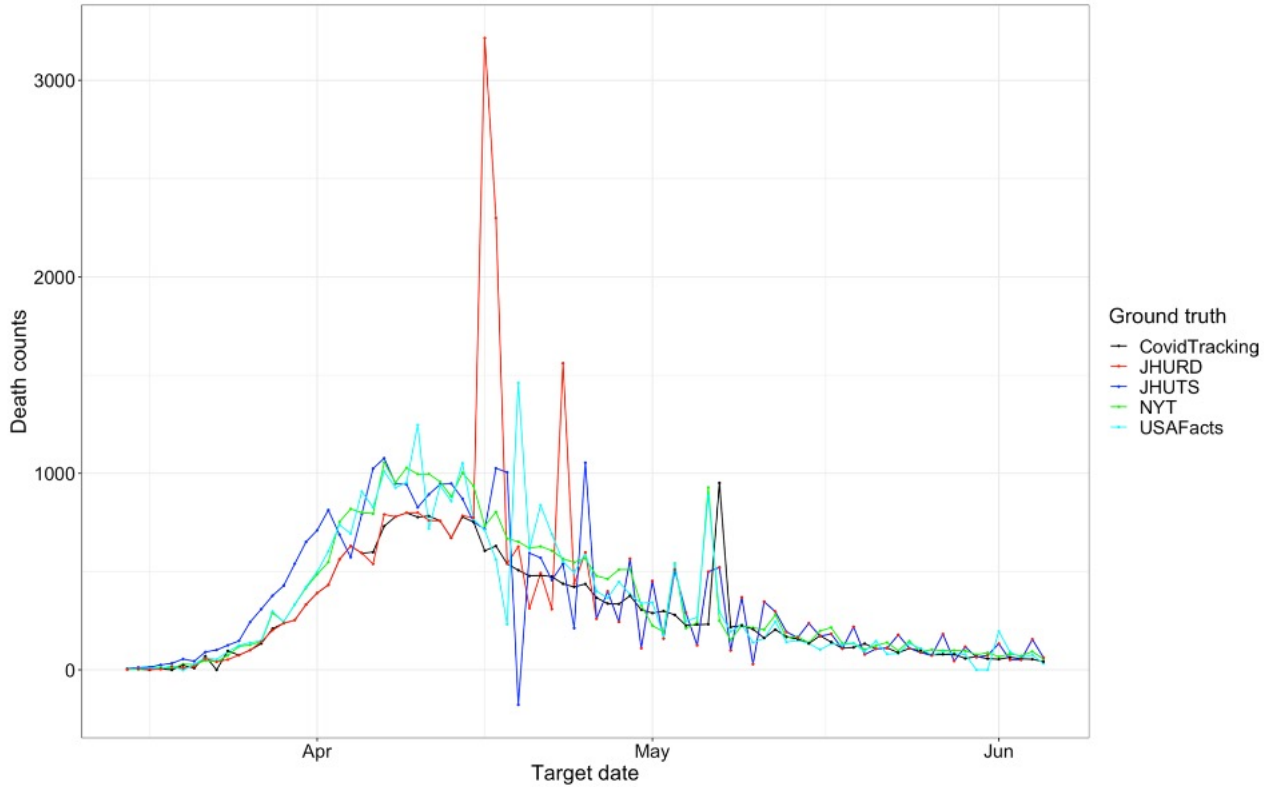


Fig. 1 A comparison of the daily death counts ground truth from CovidTracking (black), JHURD (red), JHUTS (dark blue), NYT (green) and USAFacts (light blue) for the period March 15–June 5 for NY

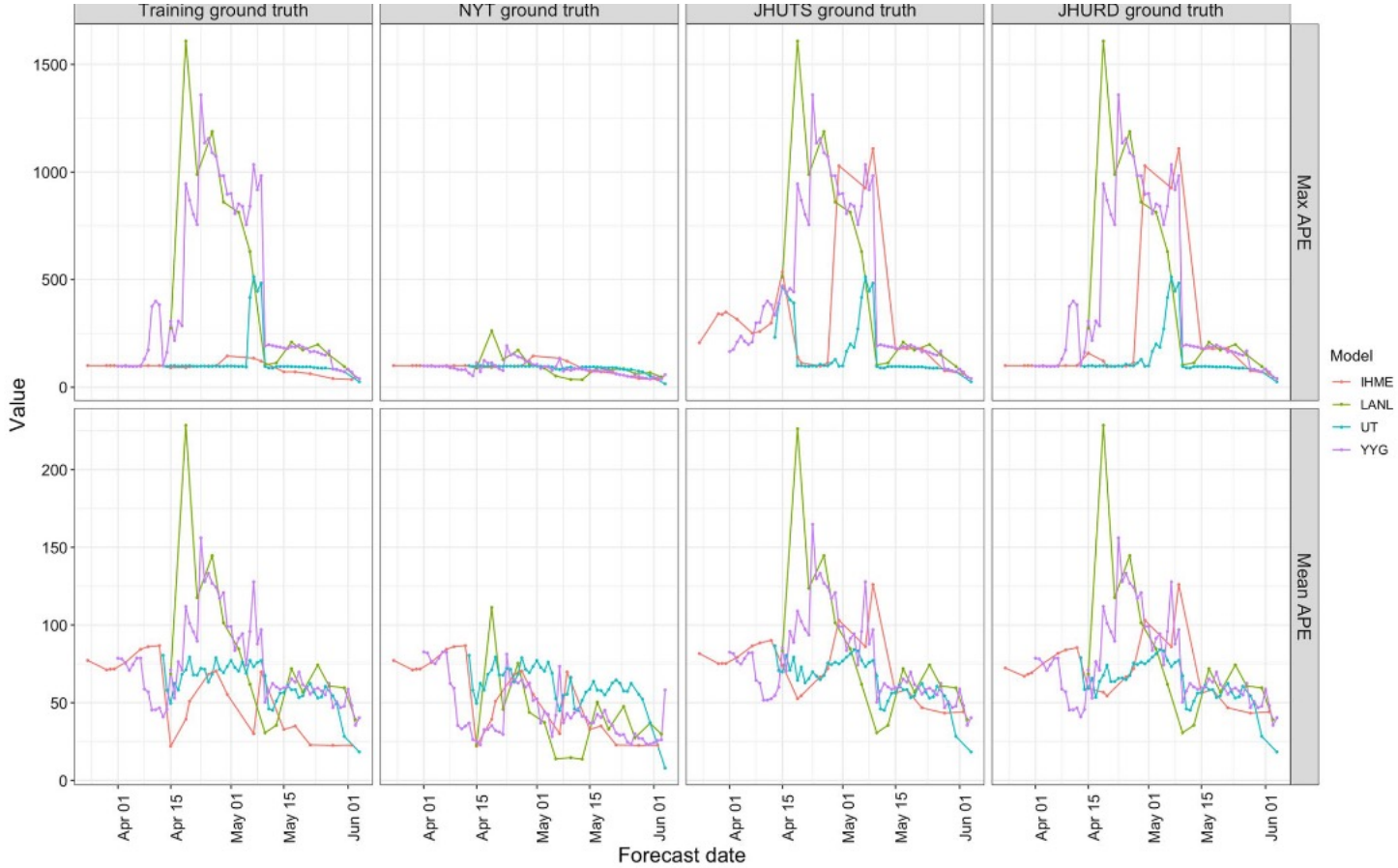


Fig. 3 Discrepancies between each model and the ground truth, as measured by the maximum absolute percentage error (top) and the mean absolute percentage error (bottom), for each version of the ground truth

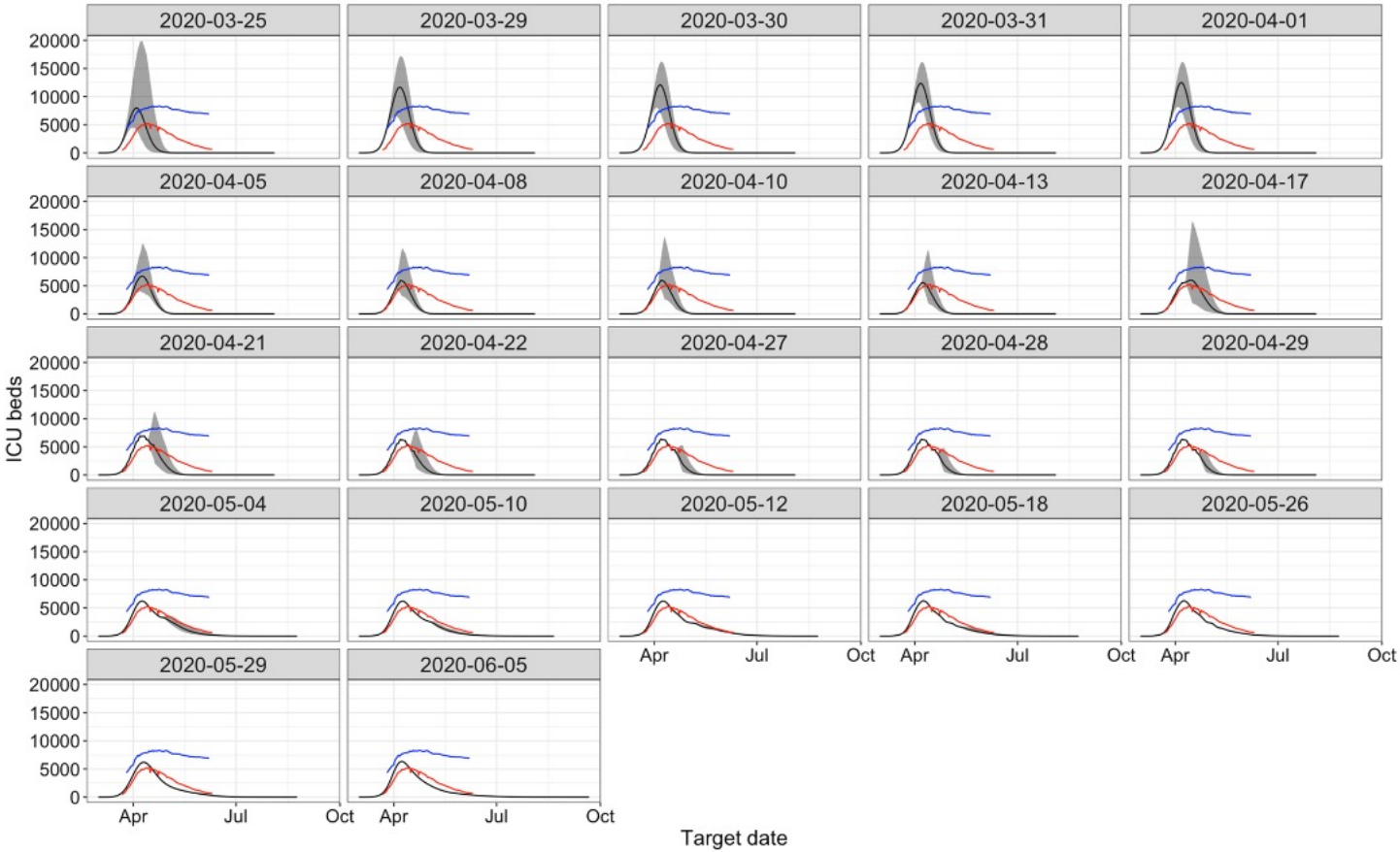


Fig. 6 Predicted ICU bed usage (black) and its 95% PIs (grey shaded area) in NY for each reporting date, along with the ground truth (red) and the maximum ICU capacity inclusive of non-COVID-19 ICU beds (blue) obtained from THE CITY

Broader considerations for (failed) forecasting in infectious diseases and pandemics

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting for COVID-19 has failed

John P.A. Ioannidis [a,*], Sally Cripps [b], Martin A. Tanner [c]

[a] *Stanford Prevention Research Center, Department of Medicine, and Departments of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, Stanford University, and Meta-Research Innovation Center at Stanford (METRICS), Stanford, CA, USA*

[b] *School of Mathematics and Statistics, The University of Sydney and Data Analytics for Resources and Environments (DARE) Australian Research Council, Sydney, Australia*

[c] *Department of Statistics, Northwestern University, Evanston, IL, USA*

Table 3
Potential reasons for the failure of COVID-19 forecasting along with examples and extent of potential amendments.

Reasons	Examples	How to fix: extent of potential amendments
Poor data input on key features of the pandemic that go into theory-based forecasting (e.g. SIR models)	Early data providing estimates for case fatality rate, infection fatality rate, basic reproductive number, and other key numbers that are essential in modeling were inflated.	May be unavoidable early in the course of the pandemic when limited data are available; should be possible to correct when additional evidence accrues about true spread of the infection, proportion of asymptomatic and non-detected cases, and risk-stratification. Investment should be made in the collection, cleaning, and curation of data.
Poor data input for data-based forecasting (e.g. time series)	Lack of consensus as to what is the 'ground truth" even for seemingly hard-core data such as the daily the number of deaths. They may vary because of reporting delays, changing definitions, data errors, etc. Different models were trained on different and possibly highly inconsistent versions of the data.	As above: investment should be made in the collection, cleaning, and curation of data.
Wrong assumptions in the modeling	Many models assume homogeneity, i.e. all people having equal chances of mixing with each other and infecting each other. This is an untenable assumption and, in reality, tremendous heterogeneity of exposures and mixing is likely to be the norm. Unless this heterogeneity is recognized, estimates of the proportion of people eventually infected before reaching herd immunity can be markedly inflated	Need to build probabilistic models that allow for more realistic assumptions; quantify uncertainty and continuously re-adjust models based on accruing evidence
High sensitivity of estimates	For models that use exponentiated variables, small errors may result in major deviations from reality	Inherently impossible to fix; can only acknowledge that uncertainty in calculations may be much larger than it seems

Lack of incorporation of epidemiological features	Almost all COVID-19 mortality models focused on number of deaths, without considering age structure and comorbidities. This can give very misleading inferences about the burden of disease in terms of quality-adjusted life-years lost, which is far more important than simple death count. For example, the Spanish flu killed young people with average age of 28 and its burden in terms of number of quality-adjusted person-years lost was about 1000-fold higher than the COVID-19 (at least as of June 8, 2020).	Incorporate best epidemiological estimates of age structure and comorbidities in the modeling; focus on quality-adjusted life-years rather than deaths
Poor past evidence on effects of available interventions	The core evidence to support "flatten-the-curve" efforts was based on observational data from the 1918 Spanish flu pandemic on 43 US cites. These data are >100-years old, of questionable quality, unadjusted for confounders, based on ecological reasoning, and pertaining to an entirely different (influenza) pathogen that had ~100-fold higher infection fatality rate than SARS-CoV-2. Even thus, the impact on reduction of total deaths was of borderline significance and very small (10%–20% relative risk reduction); conversely, many models have assumed a 25-fold reduction in deaths (e.g. from 510,000 deaths to 20,000 deaths in the Imperial College model) with adopted measures	While some interventions in the broader package of lockdown measures are likely to have beneficial effects, assuming huge benefits is incongruent with past (weak) evidence and should be avoided. Large benefits may be feasible from precise, focused measures (e.g. early, intensive testing with thorough contact tracing for the early detected cases, so as not to allow the epidemic wave to escalate [e.g. Taiwan or Singapore]; or draconian hygiene measures and thorough testing in nursing homes) rather than from blind lockdown of whole populations.

Reasons	Examples	How to fix: extent of potential amendments
Lack of transparency	The methods of many models used by policy makers were not disclosed; most models were never formally peer-reviewed, and the vast majority have not appeared in the peer-reviewed literature even many months after they shaped major policy actions	While formal peer-review and publication may unavoidably take more time, full transparency about the methods and sharing of the code and data that inform these models is indispensable. Even with peer-review, many papers may still be glaringly wrong, even in the best journals.
Errors	Complex code can be error-prone, and errors can happen even by experienced modelers; using old-fashioned software or languages can make things worse; lack of sharing code and data (or sharing them late) does not allow detecting and correcting errors	Promote data and code sharing; use up-to-date and well-vetted tools and processes that minimize the potential for error through auditing loops in the software and code
Lack of determinacy	Many models are stochastic and need to have a large number of iterations run, perhaps also with appropriate burn-in periods; superficial use may lead to different estimates	Promote data and code sharing to allow checking the use of stochastic processes and their stability
Looking at only one or a few dimensions of the problem at hand	Almost all models that had a prominent role in decision-making focused on COVID-19 outcomes, often just a single outcome or a few outcomes (e.g. deaths or hospital needs). Models prime for decision-making need to take into account the impact on multiple fronts (e.g. other aspects of health care, other diseases, dimensions of the economy, etc.)	Interdisciplinarity is desperately needed; as it is unlikely that single scientists or even teams can cover all this space, it is important for modelers from diverse ways of life to sit at the same table. Major pandemics happen rarely, and what is needed are models which combine information from a variety of sources. Information from data, from experts in the field, and from past pandemics, need to combined in a logically consistent fashion if we wish to get any sensible predictions.

Lack of expertise in crucial disciplines	The credentials of modelers are sometimes undisclosed; when they have been disclosed, these teams are led by scientists who may have strengths in some quantitative fields, but these fields may be remote from infectious diseases and clinical epidemiology; modelers may operate in subject matter vacuum	Make sure that the modelers' team is diversified and solidly grounded in terms of subject matter expertise
Groupthink and bandwagon effects	Models can be tuned to get desirable results and predictions; e.g. by changing the input of what are deemed to be plausible values for key variables. This is especially true for models that depend on theory and speculation, but even data-driven forecasting can do the same, depending on how the modeling is performed. In the presence of strong groupthink and bandwagon effects, modelers may consciously fit their predictions to what is the dominant thinking and expectations – or they may be forced to do so.	Maintain an open-minded approach; unfortunately, models are very difficult, if not impossible, to pre-register, so subjectivity is largely unavoidable and should be taken into account in deciding how much forecasting predictions can be trusted
Selective reporting	Forecasts may be more likely to be published or disseminated if they are more extreme	Very difficult to diminish, especially in charged environments; needs to be taken into account in appraising the credibility of extreme forecasts

RESEARCH ARTICLE

A meta-epidemiological assessment of transparency indicators of infectious disease models

Emmanuel A. Zavalis[1,2☯], **John P. A. Ioannidis**[1,3☯*]

1 Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, United States of America, **2** Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Solna, Stockholm, Sweden, **3** Departments of Medicine, of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, Stanford University, Stanford, California, United States of America

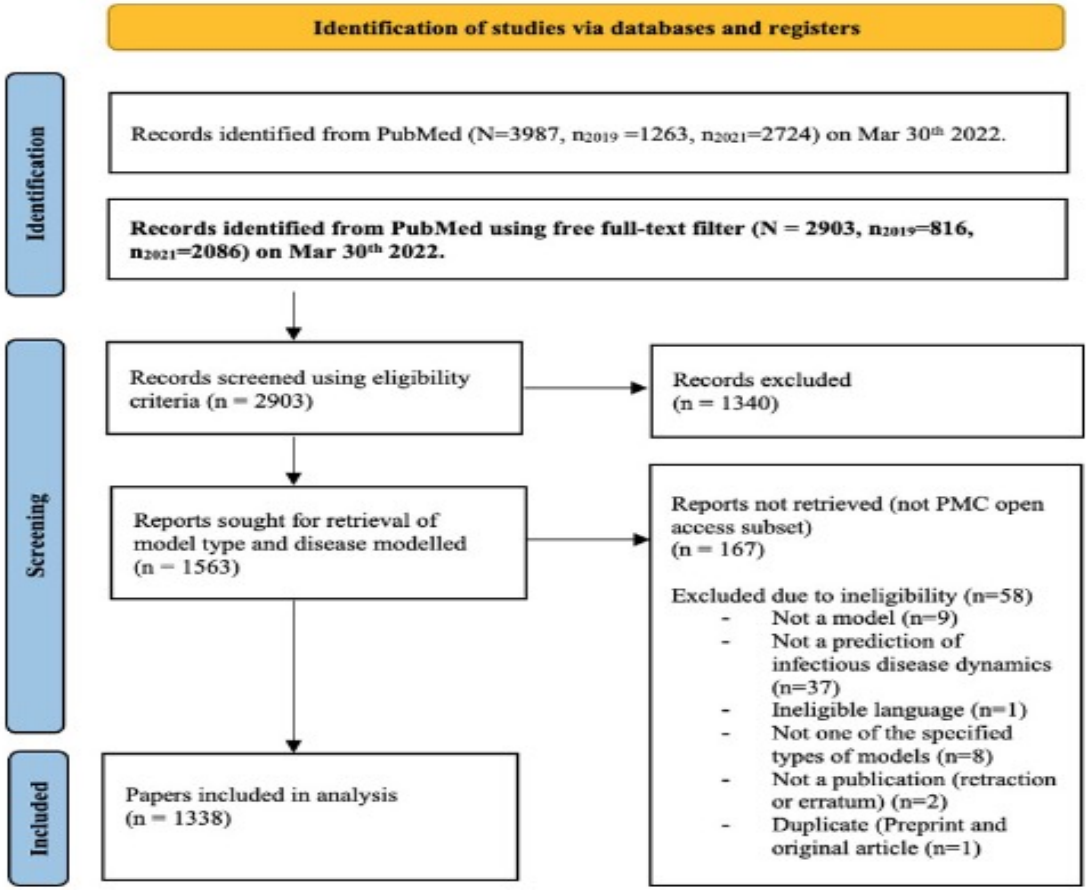


Fig 1. Flow chart for study selection.

https://doi.org/10.1371/journal.pone.0275380.g001

Table 1. Characteristics of eligible studies.

	2019	2021 non-COVID-19	2021 COVID-19	All publications
	N (%)	N (%)	N (%)	N (%)
	216 articles	304 articles	818 articles	1338 articles
Type of model				
Compartmental	26 (12.0)	91 (29.9)	394 (48.0)	511 (39.2)
Time series	80 (37.0)	82 (27.0)	175 (21.4)	337 (25.2)
Spatiotemporal	78 (36.1)	90 (29.6)	111 (13.6)	279 (20.9)
Agent-based	31 (14.4)	37 (12.2)	118 (14.4)	186 (13.9)
Multiple	1 (0.5)	4 (1.3)	20 (2.4)	25 (1.9)
Type of disease				
COVID-19	0 (0)	0 (0)	818 (100)	818 (61.1)
General	33 (15.3)	97 (31.9)	0 (0)	130 (9.7)
Influenza illnesses	20 (9.3)	20 (6.6)	0 (0)	40 (3.0)
Malaria	15 (6.9)	22 (7.2)	0 (0)	37 (2.8)
Dengue	15 (6.9)	20 (6.6)	0 (0)	35 (2.6)
Others	133 (61.6)	145 (48)	0 (0)	278 (20.8)
Journal				
PLoS One	26 (12.0)	27 (8.9)	62 (7.6)	115 (8.6)
Sci Rep	20 (9.3)	19 (6.3)	52 (6.4)	91 (6.8)
Int J Environ Res Public Health	15 (6.9)	21 (6.9)	27 (3.3)	63 (4.7)
BMC Infect Dis	16 (7.4)	12 (3.9)	10 (1.2)	38 (2.8)
PLoS Negl Trop Dis	11 (5.1)	22 (7.2)	0 (0)	33 (2.5)
PLoS Comput Biol	10 (4.6)	10 (3.3)	9 (1.1)	29 (2.2)
BMC Public Health	6 (2.8)	9 (3.0)	13 (1.6)	28 (2.1)
Chaos Solitons Fractals	0 (0)	5 (1.6)	20 (2.4)	25 (1.9)
Others	112 (52.0)	179 (58.9)	625 (76.4)	916 (68.5)

https://doi.org/10.1371/journal.pone.0275380.t001

Table 2. Key transparency indicators overall and per year/COVID-19 focus.

N = 1338	Code sharing	Data sharing	Registration	COI	Funding
	N (%)	N (%)	N (%)	N (%)	N (%)
Overall	288 (21.5)	332 (24.8)	6 (0.4)	1197 (89.5)	1109 (82.9)
2019	38 (17.6)	59 (27.3)	3 (1.4)	197 (91.2)	202 (93.5)
2021	250 (22.3)	273 (24.3)	3 (0.3)	1000 (89.2)	907 (80.8)
COVID-19	207 (25.3)	199 (24.3)	0	730 (89.2)	635 (77.6)
non-COVID-19	43 (14.1)	74 (24.3)	3 (1)	270 (88.8)	272 (89.5)
Fisher's exact test (p-values)					
2019 vs 2021	0.15	0.35	0.06	0.45	1.0×10^{-6}
2019 vs 2021 non-COVID-19	0.33	0.48	0.70	0.46	0.12
2021 non-COVID-19 vs. COVID-19	5.1×10^{-5}	1	0.02	0.83	3.5×10^{-5}

COI: conflicts of interest

https://doi.org/10.1371/journal.pone.0275380.t002

Table 3. Key transparency indicators per disease type, model type, and journal.

	Code sharing	Data sharing	Registration	COI	Funding
	N (%)	N (%)	N (%)	N (%)	N (%)
Disease modelled					
p (Fisher's exact test)	7.4×10^{-6}	0.47	0.001	0.01	2.8×10^{-10}
COVID-19	207 (25.3)	199 (24.3)	0 (0)	730 (89.2)	635 (77.6)
General (theoretical model)	31 (23.8)	34 (26.2)	0 (0)	94 (72.3)	108 (83.1)
Influenza illnesses	6 (15)	10 (25)	0 (0)	38 (95)	39 (97.5)
Malaria	2 (5.4)	7 (18.9)	2 (5.4)	37 (100)	35 (94.6)
Dengue	12 (34.3)	13 (37.1)	0 (0)	35 (100)	35 (100)
Other diseases	30 (10.8)	69 (24.8)	4 (1.4)	263 (94.6)	257 (92.4)
Type of model					
p (Fisher's exact test)	0.001	0.006	0.15	$<1 \times 10^{-7}$	0.008
Compartmental	104 (20.4)	103 (20.2)	0 (0)	419 (82)	405 (79.3)
Time Series	65 (19.3)	81 (24)	2 (0.6)	319 (94.7)	276 (81.9)
Spatiotemporal	52 (18.6)	84 (30.1)	3 (1.1)	263 (94.3)	247 (88.5)
Agent-based	63 (33.9)	58 (31.2)	1 (0.5)	173 (93)	161 (86.6)
Multiple	4 (16)	6 (24)	0 (0)	23 (92)	20 (80)
Journal					
p (Fisher exact)	0.15	1.7×10^{-12}	0.11	2.5×10^{-12}	3.4×10^{-14}
PLoS One	30 (26.1)	63 (54.8)	1 (0.9)	115 (100)	115 (100)
Sci Rep	23 (25.3)	21 (23.1)	1 (1.1)	91 (100)	70 (76.9)
Int J Environ Res Public Health	8 (12.7)	7 (11.1)	1 (1.6)	63 (100)	63 (100)
Other journals	227 (21.2)	241 (22.5)	3 (0.3)	928 (86.8)	861 (80.5)

COI: conflicts of interest

https://doi.org/10.1371/journal.pone.0275380.t003

Manual validation

We also checked a random sample of 29 (10%) of papers that were found to be sharing code, 33 (10%) of those sharing data, and all 6 that were registered. Of these, 24/29 (82.8%) actually shared code, 29/33 (87.9%) actually shared data and 5/6 (83.3%) were indeed registered. The papers that used registration were two malaria models [24, 25], one vector model [26] (which focused on malaria vectors) one polio (Sabin 2 virus [27]) model and one rotavirus model [28]. The majority were from 2021 [24, 26, 27] and were also malaria models (two malaria and one vector that was essentially malaria [24–26]). A majority was also classified as spatiotemporal [24–26]. We also checked a random sample of 10% of the negatives i.e. the ones that were classified as non-transparent and found that 133/133 (100%) weren't registered, 95/106 (89.6%) didn't share code and 75/101 (74.3%) didn't share data. Therefore, the corrected estimates of the proportions of publications sharing code and sharing data were $(0.215 \times 0.828) + (0.785 \times 0.104) = 26.0\%$ and $(0.248 \times 0.879) + (0.752 \times 0.257) = 41.1\%$, respectively. The modest number of false-negatives for detecting data sharing through the text mining algorithms reflected mostly situations where it was mentioned that the data can be downloaded through a link, or the reference was in a figure, or the phrasing was interwined and difficult to separate effectively by the text mining algorithm.

Finally, of the 120 articles (10%) that text mining found that they contained a COI statement, there was indeed a placeholder for this statement in all articles, but the vast majority of the statements (115 (95.8%)) disclosed no conflict at all. Of the 111 (10%) articles where text mining found that they contained a funding statement, all of them had indeed such a statement, but 13 (11.7%) stated that they had no funding. Examining a random sample of 10% of the negatives regarding COI and funding disclosures we found that 19/23 (82.6%) of funding disclosures and 14/14 (100%) of COI disclosures were true negatives.

ELSEVIER Journal of Clinical Epidemiology 136 (2021) 96–132

ORIGINAL ARTICLE

Effect estimates of COVID-19 non-pharmaceutical interventions are non-robust and highly model-dependent

Vincent Chin[a,b], John P.A. Ioannidis[d,e,f,g,h,*], Martin A. Tanner[c], Sally Cripps[a,b]

Abstract

Objective: To compare the inference regarding the effectiveness of the various non-pharmaceutical interventions (NPIs) for COVID-19 obtained from different SIR models.

Study design and setting: We explored two models developed by Imperial College that considered only NPIs without accounting for mobility (model 1) or only mobility (model 2), and a model accounting for the combination of mobility and NPIs (model 3). Imperial College applied models 1 and 2 to 11 European countries and to the USA, respectively. We applied these models to 14 European countries (original 11 plus another 3), over two different time horizons.

Results: While model 1 found that lockdown was the most effective measure in the original 11 countries, model 2 showed that lockdown had little or no benefit as it was typically introduced at a point when the time-varying reproduction number was already very low. Model 3 found that the simple banning of public events was beneficial, while lockdown had no consistent impact. Based on Bayesian metrics, model 2 was better supported by the data than either model 1 or model 3 for both time horizons.

Conclusion: Inferences on effects of NPIs are non-robust and highly sensitive to model specification. In the SIR modeling framework, the impacts of lockdown are uncertain and highly model-dependent.

Different models, different inferences

Chin, Ioannidis, Tanner, Cripps.
J Clin Epidemiol 2021

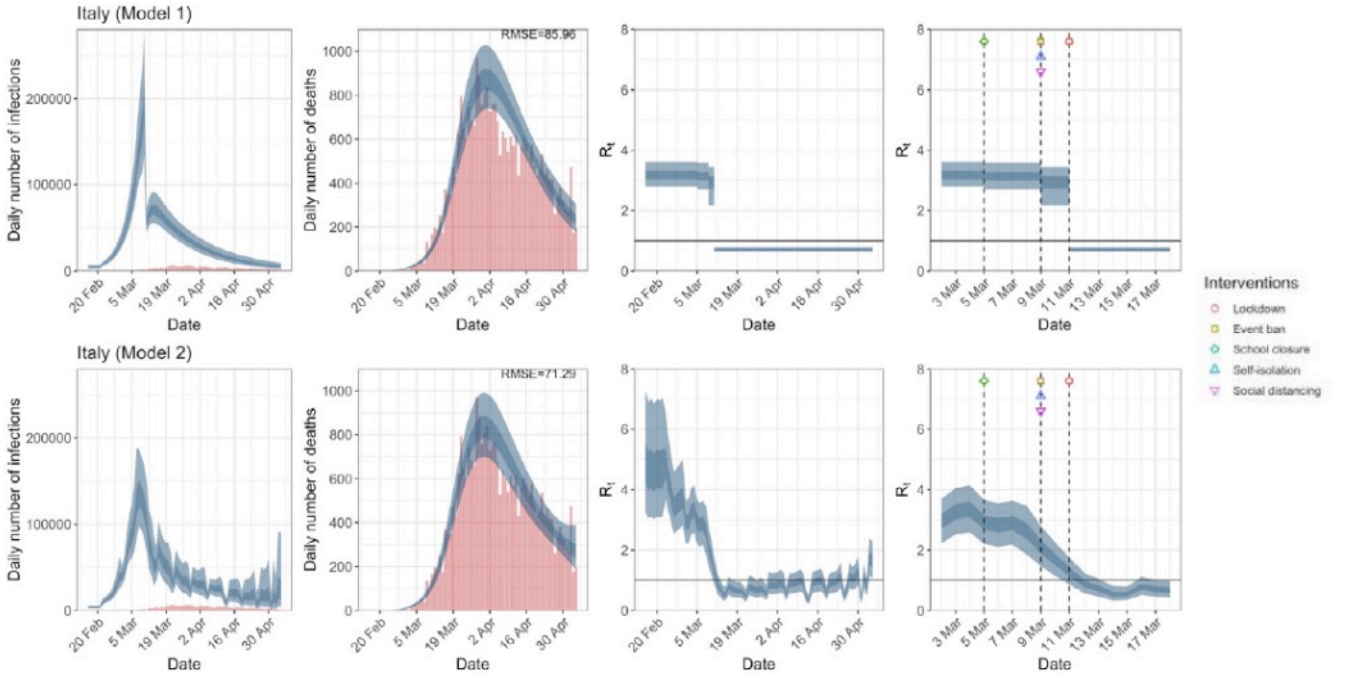


Table 1. Comparison of the value of R_t at lockdown (LD) and its 95% CIs between models 1 and 2 for all eleven countries analyzed in Flaxman et al. [1] for the time horizon March 4th to May 5th. Values of basic reproduction number R_0 and R_t immediately after the introduction of other NPIs for both models are given in Table A.5 in the Appendix.

	Model 1			Model 2
Country	R_t one day before LD	R_t at LD	% change	R_t at LD
UK	3.39	0.68	−79.67	1.11
	(2.84, 3.94)	(0.55, 0.81)	(−85.29, −72.96)	(0.75, 1.60)
Austria	2.96	0.52	−81.42	0.87
	(1.67, 4.50)	(0.40, 0.64)	(−88.80, −69.47)	(0.42, 1.55)
Belgium	4.30	0.90	−78.31	4.83
	(2.87, 6.06)	(0.78, 1.02)	(−85.99, −67.26)	(3.47, 6.45)
Denmark	3.25	0.68	−78.11	0.58
	(1.98, 4.81)	(0.57, 0.80)	(−86.01, −65.70)	(0.28, 1.05)
France	4.06	0.71	−82.08	1.69
	(2.98, 4.95)	(0.61, 0.82)	(−87.07, −74.21)	(1.16, 2.39)
Germany	3.68	0.73	−79.99	1.02
	(2.94, 4.51)	(0.60, 0.85)	(−85.84, −72.48)	(0.68, 1.47)
Italy	2.90	0.70	−75.35	1.30
	(2.17, 3.46)	(0.63, 0.78)	(−80.98, −66.51)	(0.86, 1.76)
Norway	2.42	0.40	−82.30	0.50
	(1.36, 3.71)	(0.25, 0.57)	(−91.04, −69.16)	(0.27, 0.79)
Spain	4.29	0.67	−84.05	1.78
	(3.35, 5.39)	(0.59, 0.75)	(−88.43, −78.72)	(1.22, 2.42)
Sweden	–	–	–	–
Switzerland	2.67	0.55	−78.61	0.93
	(1.93, 3.48)	(0.44, 0.68)	(−86.43, −67.32)	(0.62, 1.31)

Table 2. Estimates and standard errors of the differences of various information criteria against model 1; the Watanabe-Akaike information criterion, $WAIC1 = -2\text{lppd} + 2p_{WAIC1}$ and $WAIC2 = -2\text{lppd} + 2p_{WAIC2}$ which uses lppd as a measure of fit with p_{WAIC1} and p_{WAIC2} as the effective number of parameters to penalize the fit respectively; the Deviance information criterion DIC $= -2\log p(\mathbf{y}|\hat{\theta}_{Bayes}) + 2p_{DIC}$ which uses $\log p(\mathbf{y}|\hat{\theta}_{Bayes})$, as the measure of fit, and p_{DIC} as the penalty. Note that a negative value implies a better predictive model compared to model 1, and the preferred model for each criteria and time period is shown in bold. See Appendix B for computational details.

Model	Time period	Δ_{WAIC1}	Δ_{WAIC2}	Δ_{DIC}
2	Up to May 5th	-31.21 ± 0.30	-29.95 ± 0.34	-30.46 ± 0.28
3	Up to May 5th	-24.03 ± 0.31	-22.49 ± 0.36	-23.29 ± 0.29
2	Up to July 12th	-54.27 ± 1.78	-49.93 ± 3.42	-51.95 ± 0.37
3	Up to July 12th	-36.74 ± 1.30	-32.24 ± 3.22	-34.97 ± 0.37

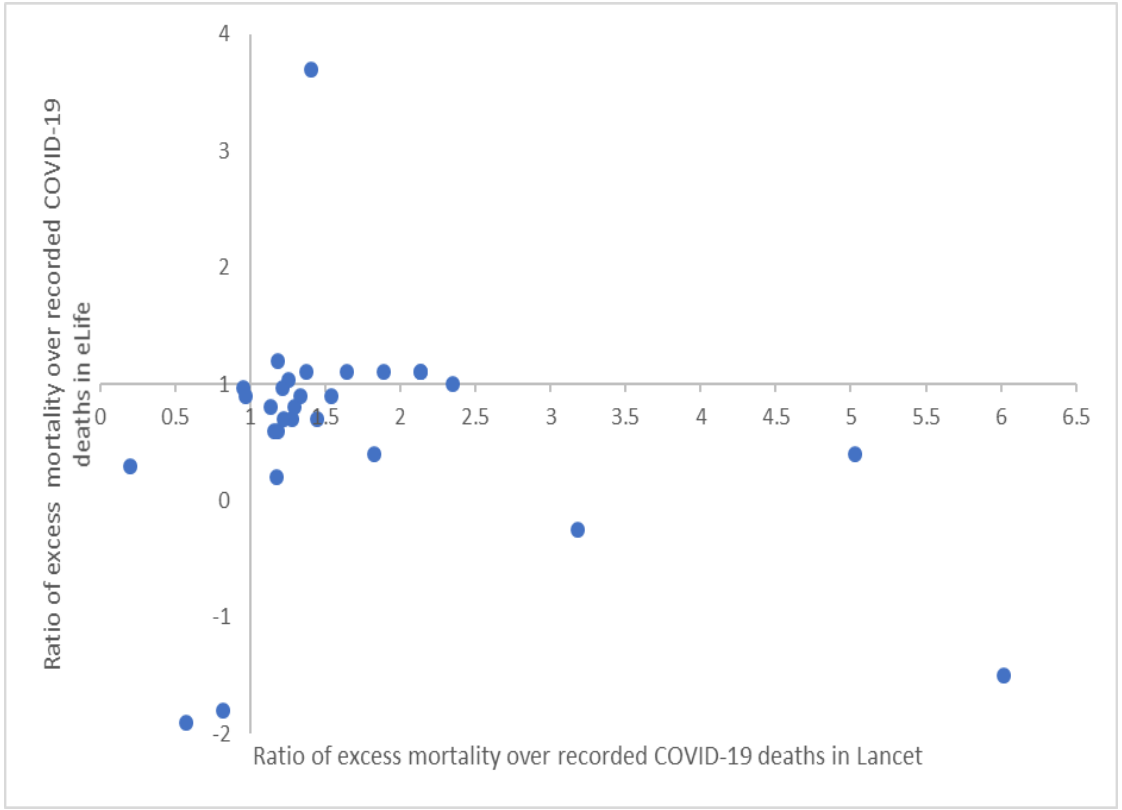
Table 3. Comparison of the value of R_t at lockdown (LD) and its 95% CIs between models 1 and 2 for all eleven countries analyzed in Flaxman et al. [1] and an additional three countries of Greece, Netherlands and Portugal, for the time horizon March 4th to July 12th.

Country	Model 1			Model 2
	R_t one day before LD	R_t at LD	% change	R_t at LD
UK	3.08	0.81	−73.25	1.20
	(2.32, 3.78)	(0.76, 0.86)	(−79.28, −64.03)	(0.72, 1.82)
Austria	1.82	0.61	−64.58	0.72
	(1.16, 2.81)	(0.55, 0.67)	(−78.02, −47.53)	(0.30, 1.42)
Belgium	2.10	0.70	−65.58	1.43
	(1.46, 2.98)	(0.67, 0.73)	(−76.83, −51.27)	(0.90, 2.05)
Denmark	1.73	0.68	−59.12	0.56
	(1.16, 2.48)	(0.60, 0.76)	(−72.79, −41.89)	(0.25, 1.05)
France	2.26	0.71	−67.37	1.77
	(1.59, 3.12)	(0.67, 0.75)	(−77.65, −53.86)	(1.11, 2.60)
Germany	3.31	0.71	−78.13	1.12
	(2.51, 4.19)	(0.66, 0.76)	(−83.73, −70.87)	(0.69, 1.67)
Italy	1.74	0.75	−55.66	1.41
	(1.26, 2.32)	(0.71, 0.79)	(−68.31, −39.35)	(0.88, 2.03)
Norway	1.52	0.57	−60.72	0.53
	(0.97, 2.22)	(0.48, 0.66)	(−74.83, −40.59)	(0.27, 0.88)
Spain	3.47	0.75	−77.74	1.74
	(2.51, 4.46)	(0.72, 0.79)	(−83.34, −69.56)	(1.07, 2.49)
Sweden	–	–	–	–
Switzerland	1.76	0.61	−64.49	0.96
	(1.25, 2.41)	(0.57, 0.64)	(−75.75, −50.23)	(0.58, 1.39)
Greece	1.46	0.69	−51.03	0.35
	(0.90, 2.05)	(0.63, 0.74)	(−67.21, −22.64)	(0.16, 0.61)
Netherlands	1.77	0.66	−62.14	1.00
	(1.34, 2.25)	(0.61, 0.70)	(−72.27, −49.34)	(0.61, 1.44)
Portugal	1.74	0.83	−50.31	0.66
	(1.12, 2.39)	(0.80, 0.86)	(−65.50, −25.24)	(0.36, 1.07)

Table A.4. RMSE of daily death counts for models 1 and 2 for the data up to May 5th and July 12th. A lower RMSE between models 1 and 2 for each country is shown in bold.

Country	Up to May 5th		Up to July 12th	
	Model 1	Model 2	Model 1	Model 2
UK	**145.41**	145.64	134.26	**129.68**
Austria	**5.88**	**5.88**	**4.48**	4.57
Belgium	71.16	**52.91**	25.20	**15.84**
Denmark	3.27	**3.08**	2.42	**2.39**
France	242.07	**227.22**	187.33	**168.34**
Germany	**48.62**	48.75	37.04	**36.32**
Italy	85.96	**71.29**	63.47	**57.42**
Norway	**3.06**	3.07	**2.21**	2.22
Spain	95.23	**92.43**	143.82	**135.03**
Sweden	35.82	**35.55**	33.12	**33.09**
Switzerland	14.61	**14.34**	10.37	**10.31**
Greece			1.72	**1.51**
Netherlands			21.48	**21.01**
Portugal			6.29	**5.75**

Excess deaths in high-income countries per Lancet: comparison with eLife



Environmental Research 213 (2022) 113754

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Comparison of pandemic excess mortality in 2020–2021 across different empirical calculations

Michael Levitt [a], Francesco Zonta [b], John P.A. Ioannidis [c,d,e,f,g,*]

Excess death calculations depend on modeling (Levitt, Zonta, Ioannidis, Envir Res 2022)

We assessed excess deaths for the entire two-year period 2020-2021 33 high-income countries with available weekly mortality data according to age strata in mortality.org.
Total population of 1 billion, **1.9 million** COVID19 deaths recorded during this period.

Three published modeling calculations do not use age-adjustment
The eLife modeling estimates **2.0 million** deaths
The Economist modeling estimates **2.2 million** deaths
The Lancet/IHME modeling estimates **2.8 million** deaths
Our modeling estimates **2.2 million deaths without age adjustment**
1.5 million deaths with age adjustment

A close look at excess deaths (2020-2021) in Germany

- Our age-adjusted estimate is **43,000** excess deaths
- Without age-adjustment we calculated **117,000** excess deaths
- Lancet calculated **203,000** excess deaths
- eLife calculated **88,000** excess deaths
- Economist calculated **113,000** excess deaths
- Baum (2022) calculated **22,000** excess deaths after age adjustment
- Koenig et al (2022) calculated **~130,000** excess deaths without age adjustment
- The recorded COVID-19 deaths were **111,000**

- **In Germany, the number of people aged >80 years increased from 4.8 million in 2016 to 5.8 million in 2020, so consideration of age is crucial.**

Excess death estimates from multiverse analysis in 2009-2021

Michael Levitt,[a*] Francesco Zonta,[b] John P.A. Ioannidis[c]

Table 1: Average, standard deviation, minimum, maximum and range for estimates of relative excess deaths (expressed as percentage of expected deaths, p%) for the two-year pandemic period 2020-2021 for each of the 33 countries.

Country	Abbreviation	Average p%	SD of p%	Minimum p%	Maximum p%	Range of p%
Australia	AUS	-9.7	3.2	-16.2	-2.4	13.9
Austria	AUT	3.2	3.0	-3.4	9.2	12.6
Belgium	BEL	1.4	2.9	-5.0	8.8	13.8
Canada	CAN	2.2	2.0	-4.9	6.9	11.7
Switzerland	CHE	-1.3	3.1	-8.2	5.7	13.9
Chile	CHL	6.4	3.8	-1.7	15.1	16.8
Czechia	CZE	8.7	3.9	-0.5	16.7	17.2
Germany	DEU	1.0	1.9	-4.4	4.5	8.9
Denmark	DNK	-7.6	4.0	-18.6	-0.3	18.3
Spain	ESP	3.6	2.2	-2.6	10.9	13.5
Estonia	EST	1.7	4.8	-10.8	11.0	21.9
Europe	EUM	2.3	2.2	-3.7	7.4	11.1
Finland	FIN	-5.3	3.1	-11.9	1.6	13.4
France	FRA	2.4	2.0	-3.8	6.1	10.0
United Kingdom	GBR	4.2	1.9	-1.2	10.0	11.3
Greece	GRC	5.6	2.8	-1.3	10.6	12.0
Croatia	HRV	7.0	3.1	-1.2	14.8	16.1
Hungary	HUN	6.8	2.7	0.5	13.1	12.6
Iceland	ISL	-7.3	2.0	-12.2	-2.1	10.1
Israel	ISR	-1.5	2.9	-7.1	4.6	11.6
Italy	ITA	5.4	2.4	-0.5	10.8	11.2
South Korea	KOR	-13.5	5.2	-24.5	-1.1	23.5
Lithuania	LTU	8.6	3.2	2.0	18.8	16.8
Luxembourg	LUX	-2.6	3.9	-10.6	4.4	15.0
Latvia	LVA	7.0	3.1	-1.0	14.0	15.0
Netherlands	NLD	2.5	2.0	-2.5	7.8	10.4
Norway	NOR	-9.4	3.6	-16.1	-1.4	14.7
New Zealand	NZL	-9.1	2.5	-15.5	-4.2	11.3
Poland	POL	14.2	3.5	3.9	19.9	15.9
Portugal	PRT	3.6	2.4	-2.7	8.6	11.3
Slovakia	SVK	10.2	4.4	0.7	20.7	20.0
Slovenia	SVN	4.7	3.4	-4.0	11.8	15.7
Sweden	SWE	-6.7	3.4	-12.5	4.2	16.7
United States	USA	16.6	0.8	14.3	18.7	4.3

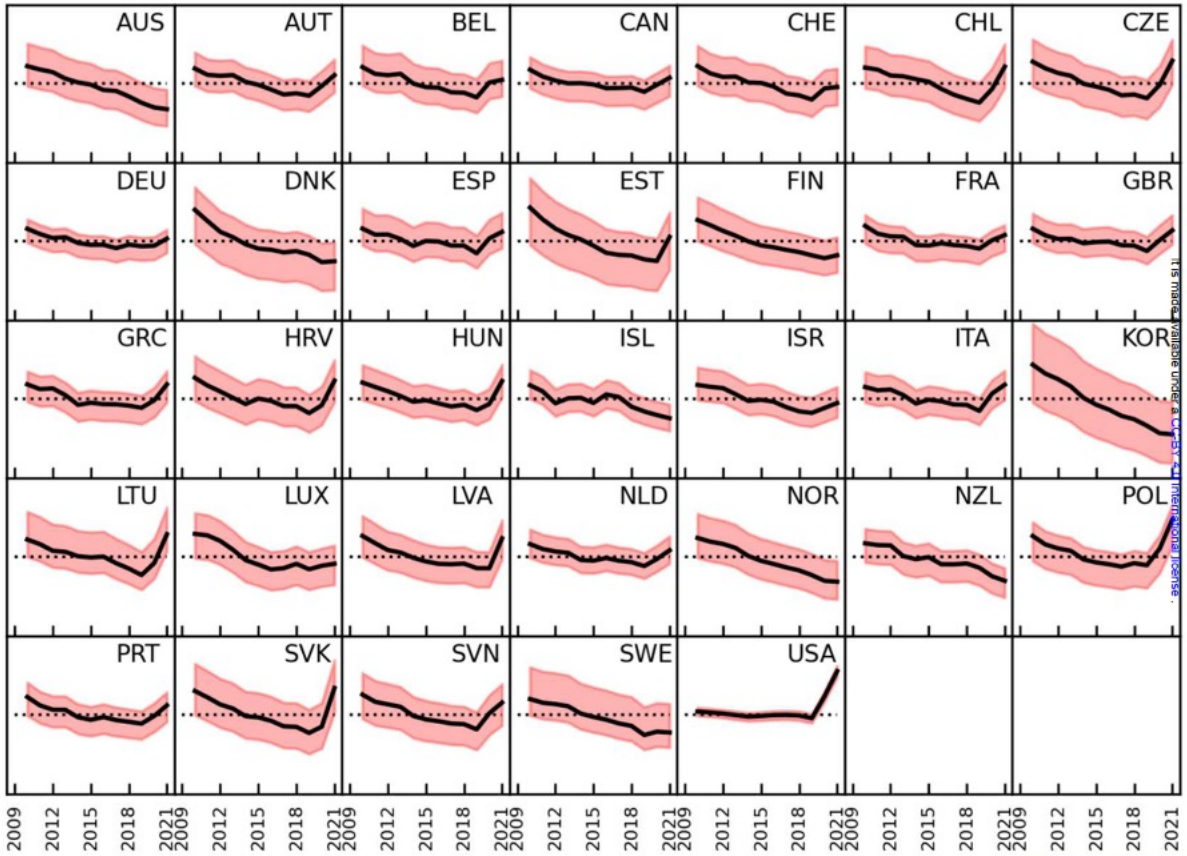

AUS
AUT
BEL
CAN
CHE
CHL
CZE
DEU
DNK
ESP
EST
FIN
FRA
GBR
GRC
HRV
HUN
ISL
ISR
ITA
KOR
LTU
LUX
LVA
NLD
NOR
NZL
POL
PRT
SVK
SVN
SWE
USA
2009
2012
2015
2018
2021
Year Between 2009 and 2021 (Summing Two Adjacent Years)

Figure 1

Table 2: Distribution of the country rank of the excess death estimates in the pandemic 2-year period 2020-2021 expressed as a percentage of the expected deaths p% for the 33 countries as calculated for each of the 66 different reference baseline year sets..

Location	LOC	Rank AVE	Rank SD	Rank in Sort from Highest p% to Lowest p% Summing Two Adjacent Years																												
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
United States	USA	1.29	0.54	49	16		1																											
Poland	POL	1.79	0.44	15	50	1																												
Slovakia	SVK	3.26	0.88	2		50	11	1		2																								
Lithuania	LTU	4.68	1.23			12	16	26	9		2	1																						
Czechia	CZE	4.73	1.05				34	23	6	1	1		1																					
Latvia	LVA	7.14	1.68				1	7	19	15	15	3	3	1	1	1																		
Croatia	HRV	7.21	1.57			1	1	4	12	27	10	6	3	1		1																		
Hungary	HUN	7.68	1.14					2	10	13	25	14	2																					
Chile	CHL	9.09	2.76			1	1	2	10	5	7	13	7	9	7	1	1			1	1													
Greece	GRC	10.23	1.84					1		1	4	16	22	10	7	2	1	1			1													
Italy	ITA	10.55	2.01			1	1			2	2	7	15	23	8	4	1	1	1															
Slovenia	SVN	12.56	2.04										10	10	23	4	4	11	1	1	1													
United Kingdom	GBR	13.08	2.02									5	1	8	10	18	5	9	9	1														
Portugal	PRT	14.62	1.64											1	2	9	26	14	6	6		1	1											
Spain	ESP	14.86	1.99									1		2	2	12	9	19	6	11	2		1											
Austria	AUT	16.35	2.51												2	8	7	7	20	2	1	11	2											
Netherlands	NLD	17.97	1.83													1	1	1	8	17	17	4	5	1		1								
France	FRA	18.52	2.11														2	1	8	12	15	8	5	9										
Estonia	EST	19.11	4.54										2	1	3	5	6	1	2	7		3	6	6	16	2	4		1					
Canada	CAN	19.38	2.37												1		3	1	4	4	5	15	13	10		1								
Belgium	BEL	21.12	1.85																1	1	11	2	8	31	10	1								
Germany	DEU	22.17	1.38																		1	2	24	8	24	3	3	1						
Switzerland	CHE	24.35	0.93																						10	33	14	8	1					
Israel	ISR	24.76	0.85																					1	5	12	40	7	1					
Luxembourg	LUX	25.62	0.93																						1	12	4	43	6					
Finland	FIN	27.29	0.54																									1	47	16	2			
Sweden	SWE	28.95	1.09																								1		2	21	22	18	1		
Iceland	ISL	29.85	2.48																							1		4	6	11	19	10	7	4	5	2
Denmark	DNK	29.88	1.75																											17	19	5	5	2	4	
New Zealand	NZL	32.20	1.46																											2		3	23	9	6	6
Norway	NOR	32.39	1.31																												8	8	25	14		
Australia	AUS	33.21	0.98																											1	1	10	23	31		
South Korea	KOR	34.80	0.70																												1		1	6	32	

Table 3: Effect of changing the width of the projected period of interest from 1 to 4 years for the most recent years (2021 alone, 2020 alone, 2020-2021, 2019-2021, 2018-2021).

		1 Year	1 Year	<2 Years>	<3 Years>	<4 Years>			
Location	LOC	2020	2021	2020 +2021	2019 +2020 +2021	2018 +2019 +2020 +2021	max (2020,2021) - <2 Years>	max (2020,2021) - <3 Years>	max (2020,2021) - <4 Years>
Australia	AUS	-10.7	-8.2	-9.5	-8.7	-8.4	1.3	0.5	0.2
Austria	AUT	3.7	2.8	3.3	0.5	-0.6	0.4	3.3	4.3
Belgium	BEL	7.6	-4.7	1.4	-1.3	-1.9	6.2	8.9	9.5
Canada	CAN	3.5	1.0	2.3	0.2	-0.3	1.2	3.3	3.7
Switzerland	CHE	3.0	-5.3	-1.2	-3.0	-3.6	4.2	6.0	6.6
Chile	CHL	3.6	10.2	6.7	2.2	-0.1	3.4	8.0	10.2
Czechia	CZE	5.5	11.6	8.6	3.6	1.7	3.0	8.0	9.9
Germany	DEU	-0.2	2.1	1.0	-0.3	-0.2	1.1	2.3	2.3
Denmark	DNK	-9.0	-6.9	-7.8	-7.4	-6.1	0.9	0.5	-0.8
Spain	ESP	8.8	-1.4	3.6	0.2	-0.4	5.2	8.5	9.2
Estonia	EST	-6.8	9.4	1.5	-1.7	-2.5	7.9	11.1	11.9
Finland	FIN	-6.1	-4.6	-5.4	-5.8	-5.3	0.8	1.2	0.6
France	FRA	3.8	0.6	2.3	0.5	-0.1	1.5	3.3	3.9
United Kingdom	GBR	6.2	2.3	4.2	1.1	0.4	2.0	5.0	5.8
Greece	GRC	1.0	9.9	5.6	3.1	1.4	4.3	6.8	8.6
Croatia	HRV	1.9	11.8	6.9	2.3	0.9	4.9	9.5	10.9
Hungary	HUN	1.6	11.8	6.7	2.7	1.4	5.1	9.1	10.4
Iceland	ISL	-6.9	-7.5	-7.2	-6.5	-6.0	0.3	-0.4	-0.9
Israel	ISR	-1.9	-0.8	-1.3	-2.5	-3.3	0.5	1.6	2.4
Italy	ITA	8.9	2.1	5.5	2.0	0.5	3.4	6.9	8.3
South Korea	KOR	-13.1	-13.0	-13.1	-12.8	-11.6	0.2	-0.2	-1.4
Lithuania	LTU	3.9	13.3	8.5	2.7	0.7	4.9	10.6	12.7
Luxembourg	LUX	-0.6	-4.8	-2.7	-3.8	-3.4	2.2	3.2	2.8
Latvia	LVA	-2.8	16.3	6.8	2.6	1.5	9.5	13.7	14.9
Netherlands	NLD	3.3	1.8	2.5	0.1	-0.4	0.8	3.2	3.6
Norway	NOR	-10.1	-8.6	-9.4	-8.9	-8.2	0.8	0.3	-0.5
New Zealand	NZL	-10.7	-7.5	-9.0	-7.3	-6.5	1.5	-0.2	-1.0
Poland	POL	10.2	17.8	14.2	8.2	5.9	3.6	9.5	11.9
Portugal	PRT	3.6	3.3	3.5	0.9	0.2	0.1	2.7	3.4
Slovakia	SVK	-0.3	20.7	10.2	4.1	2.0	10.5	16.6	18.7
Slovenia	SVN	7.5	1.8	4.7	1.1	-0.2	2.8	6.4	7.8
Sweden	SWE	-2.2	-10.6	-6.6	-7.9	-7.2	4.5	5.7	5.1
United States	USA	15.8	17.6	16.7	10.6	7.8	0.9	7.0	9.8

Over- or under-estimation of COVID-19 deaths?

(Ioannidis, Eur J Epidemiol 2021)

X	F=0.05%, m=0.2%	F=0.1%, m=0.2%	F=0.4%, m=0.2%	F=1.0%, m=0.2%	F=0.05%, m=0.9%	F=0.1%, m=0.9%	F=0.4%, m=0.9%	F=1.0%, m=0.9%	F=0.05%, m=1.5%	F=0.1%, m=1.5%	F=0.4%, m=1.5%	F=1.0%, m=1.5%
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.72	0.36	0.09	0.04	3.24	1.62	0.41	0.16	5.40	2.70	0.68	0.27
0.2	1.44	0.72	0.18	0.07	6.48	3.24	0.81	0.32	10.80	5.40	1.35	0.54
0.3	2.16	1.08	0.27	0.11	9.72	4.86	1.22	0.49	16.20	8.10	2.03	0.81
0.4	2.88	1.44	0.36	0.14	12.96	6.48	1.62	0.65	21.60	10.80	2.70	1.08
0.5	3.60	1.80	0.45	0.18	16.20	8.10	2.03	0.81	27.00	13.50	3.38	1.35
0.6	4.32	2.16	0.54	0.22	19.44	9.72	2.43	0.97	32.40	16.20	4.05	1.62
0.7	5.04	2.52	0.63	0.25	22.68	11.34	2.84	1.13	37.80	18.90	4.73	1.89
0.8	5.76	2.88	0.72	0.29	25.92	12.96	3.24	1.30	43.20	21.60	5.40	2.16
0.9	6.48	3.24	0.81	0.32	29.16	14.58	3.65	1.46	48.60	24.30	6.08	2.43
1	7.20	3.60	0.90	0.36	32.40	16.20	4.05	1.62	54.00	27.00	6.75	2.70
1.1	7.92	3.96	0.99	0.40	35.64	17.82	4.46	1.78	59.40	29.70	7.43	2.97
1.2	8.64	4.32	1.08	0.43	38.88	19.44	4.86	1.94	64.80	32.40	8.10	3.24
1.3	9.36	4.68	1.17	0.47	42.12	21.06	5.27	2.11	70.20	35.10	8.78	3.51
1.4	10.08	5.04	1.26	0.50	45.36	22.68	5.67	2.27	75.60	37.80	9.45	3.78
1.5	10.80	5.40	1.35	0.54	48.60	24.30	6.08	2.43	81.00	40.50	10.13	4.05
1.6	11.52	5.76	1.44	0.58	51.84	25.92	6.48	2.59	86.40	43.20	10.80	4.32
1.7	12.24	6.12	1.53	0.61	55.08	27.54	6.89	2.75	91.80	45.90	11.48	4.59
1.8	12.96	6.48	1.62	0.65	58.32	29.16	7.29	2.92	97.20	48.60	12.15	4.86
1.9	13.68	6.84	1.71	0.68	61.56	30.78	7.70	3.08	102.60	51.30	12.83	5.13
2	14.40	7.20	1.80	0.72	64.80	32.40	8.10	3.24	108.00	54.00	13.50	5.40

Preserving equipoise and performing randomised trials for COVID-19 social distancing interventions

Ioana Alina Cristea[1], Florian Naudet[2] and John P. A. Ioannidis[3]

[1]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; [2]University Rennes, CHU Rennes, Inserm, CIC 1414 (Centre d'Investigation Clinique de Rennes), F-35000, Rennes, France and [3]Departments of Medicine, of Epidemiology and Population Health of Biomedical Data Science, and of Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, USA

Abstract

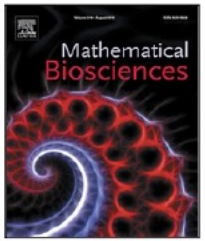
In the coronavirus disease 2019 (COVID-19) pandemic, a large number of non-pharmaceutical measures that pertain to the wider group of social distancing interventions (e.g. public gathering bans, closures of schools, workplaces and all but essential business, mandatory stay-at-home policies, travel restrictions, border closures and others) have been deployed. Their urgent deployment was defended with modelling and observational data of spurious credibility. There is major debate on whether these measures are effective and there is also uncertainty about the magnitude of the harms that these measures might induce. Given that there is equipoise for how, when and if specific social distancing interventions for COVID-19 should be applied and removed/modified during reopening, we argue that informative randomised-controlled trials are needed. Only a few such randomised trials have already been conducted, but the ones done to-date demonstrate that a randomised trials agenda is feasible. We discuss here issues of study design choice, selection of comparators (intervention and controls), choice of outcomes and additional considerations for the conduct of such trials. We also discuss and refute common counter-arguments against the conduct of such trials.

Mathematical Biosciences 345 (2022) 108782

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs



Perspective

Pre-registration of mathematical models

John P.A. Ioannidis *

Departments of Medicine, of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

ARTICLE INFO

Keywords:
Mathematical modeling
Pre-registration
Bias
Reproducibility

ABSTRACT

Pre-registration is a research practice where a protocol is deposited in a repository before a scientific project is performed. The protocol may be publicly visible immediately upon deposition or it may remain hidden until the work is completed/published. It may include the analysis plan, outcomes, and/or information about how evaluation of performance (e.g. forecasting ability) will be made, Pre-registration aims to enhance the trust one can put on scientific work. Deviations from the original plan, may still often be desirable, but pre-registration makes them transparent. While pre-registration has been advocated and used to variable extent in diverse types of research, there has been relatively little attention given to the possibility of pre-registration for mathematical modeling studies. Feasibility of pre-registration depends on the type of modeling and the ability

Table 1
Conditions that may favor or disfavor pre-registration.

Pre-registration favored	Pre-registration disfavored
Rigorous design thought in advance	Design to be fine-tuned iteratively
Standardized procedures preconceived	Procedures to be discovered
Optimal choices conceived in advance	Optimal choices unknown
Confirmatory research	Exploratory discovery research
Outcome/performance evaluation, e.g. forecasting	No outcome/performance evaluation
Projects can be separated into specific steps	Projects too chaotic even to specify steps
Data are to be collected prospectively	Existing data are used

Table 2
Potential advantages and disadvantages of pre-registration.

Potential advantages
Increased trust in research work
More objective assessment of model performance
Decrease in the possibility of bias/manipulation of results and inferences
Making research visible in public earlier
Reduction of redundancy in research efforts, better overall research agenda
Allowing to claim early credit for scientific work and ideas[a]
Potential disadvantages
Extra work needed
Fake pre-registration (registration has happened after the study was done)
Over-optimism that quality and efficiency of research would improve

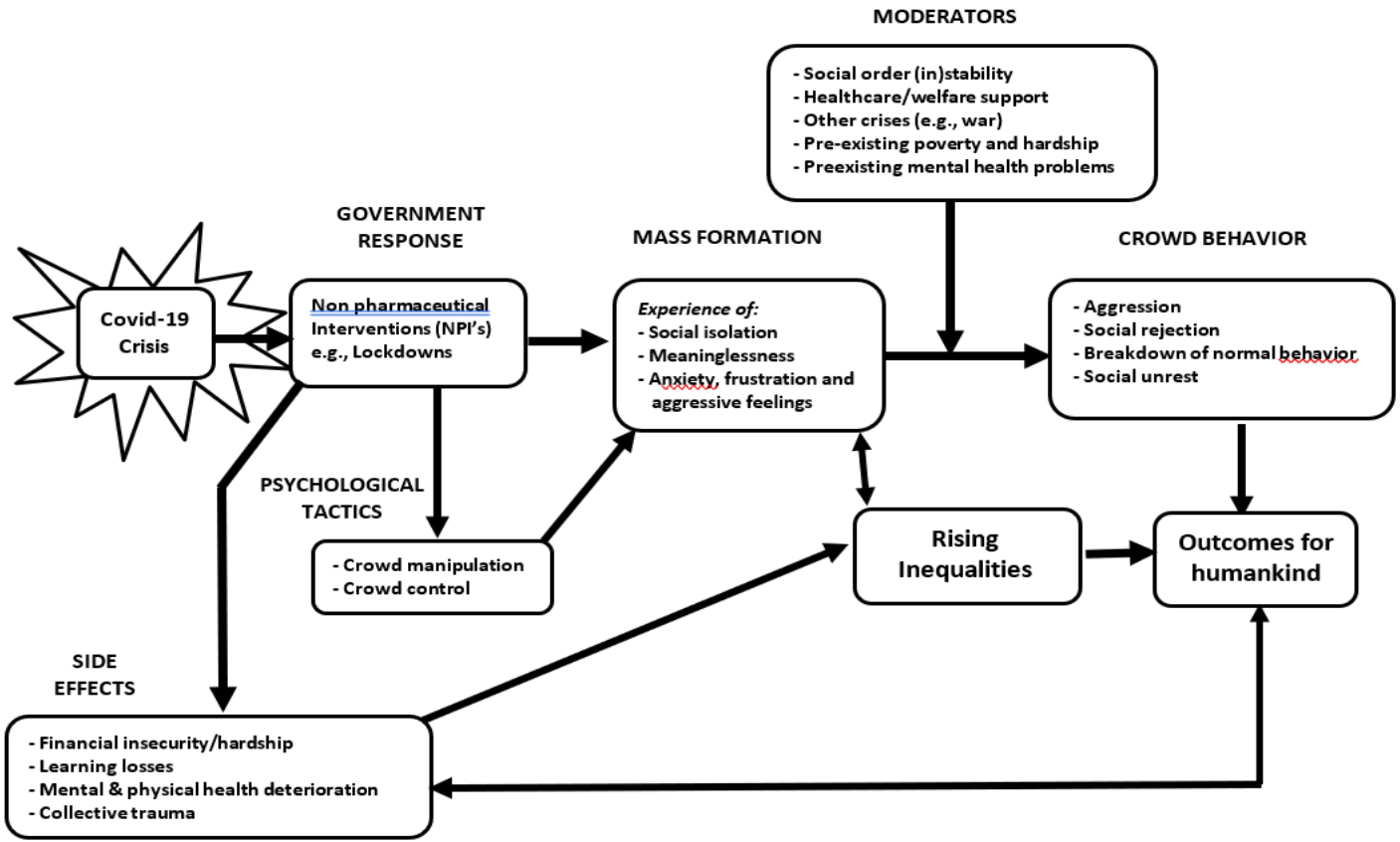
Decision-making (personal and public)

must be multi-dimensional

Ioannidis, Eur J Clin Invest 2020

Cause of excess death	Reason/comments	Possible time horizon for excess deaths
People with AMI and other acute disease not given proper hospital care	Patients afraid to go to hospital and hospitals reducing admissions afraid of overload	Acute, during pandemic
People with cancer having delayed treatment	Postponement of cancer treatment in anticipation of COVID-19 overload	Next 5 y
Disrupted cancer prevention	Inability to offer cancer prevention services under aggressive measures	Next 20 y
Other healthcare disruption	Postponement or cancellation of elective procedures and regular care	Variable for different medical conditions
Suicides	Mental health disruption	Both acute and long-term
Violence (domestic, homicide)	Mental health disruption	Acute, possibly long-term
Starvation	Disruption in food production and transport	Acute, and possibly worse over next several years
Tuberculosis	Disruption of tuberculosis management programmes	Next 5 y
Childhood diseases	Disruption of vaccination programmes	Next 5 y
Alcoholism and other diseases of despair	Mental health disruption, unemployment	Next 10 y
Multiple chronic diseases	Unemployment, lack of health insurance and poverty	Next 20 y
Lack of proper medical care	Disruption of healthcare, as hospitals and health programmes get financially disrupted, furlough personnel or even shut down services	Next 20 y

Mass formation, inequalities and long-term adverse outcomes (Schippers, Ioannidis, Joffe, Frontiers in Public Health 2022)



Some concluding comments

- Models are here to stay and they can be valuable
- Improvements are possible at the level of data input, transparency, relevance, real-life value, pre-registration (when applicable)
- Meta-epidemiological assessments can offer an observatory of how models perform and also form a basis for possible interventions to further improve them

Many thanks

- Cathrine Axfors, Stanford
- Ellen Kuhl, Stanford
- Lars Hemkens, U Basel
- Sally Cripps, U Sydney
- Martin Tanner, Northwestern
- Eran Bendavid, Stanford
- Prabhat Jha, U Toronto
- Jay Bhattacharya, Stanford
- Perrine Janiaud, Stanford and U Basel
- Michael Levitt, Stanford
- Ulrich Dirnagl, Charite and BIH
- Lu Tian, Stanford
- Bryan Cai, Stanford
- Vincent Chin, U Sydney
- David Allison, U Indiana
- Stefan Pilz, U Graz
- Roberta Pastorino, U Rome
- Marcel Ballin, U Umea
- Peter Nordstrom, U Umea
- Ioana Cristea, U Pavia
- Michael Ewers, LMU Munich
- Michaele Schippers, Erasmus U Rotterdam
- Nicolaus Plesnila, LMU Munich
- Florian Naudet, U Rennes
- Steve Goodman, Stanford
- Janneke van't Hooft, Stanford and Amsterdam
- Despina Contopoulos-Ioannidis, Stanford
- David Grimes, Queen's U Belfast
- Stefania Boccia, U Rome
- Walter Ricciardi, U Rome
- Valentin Danchev, U Essex
- Mario Malicki, Stanford
- Neeraj Sood, U Southern California
- Matthias Peirlinck, Stanford and U Ghent
- Noah Haber, Stanford
- Michael Levitt, Stanford
- Angelo Pezzullo, U Rome
- Snezana Medic, Public Health Voivodina
- Athanasios Tsakris, U Athens
- Emmanuel Zavalis, Karolinska