



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



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



ARTICLE INFO

Keywords:

Forecasting
 COVID-19
 Mortality
 Hospital bed utilization
 Bayesian models
 SIR models
 Bias
 Validation

ABSTRACT

Epidemic forecasting has a dubious track-record, and its failures became more prominent with COVID-19. Poor data input, wrong modeling assumptions, high sensitivity of estimates, lack of incorporation of epidemiological features, poor past evidence on effects of available interventions, lack of transparency, errors, lack of determinacy, consideration of only one or a few dimensions of the problem at hand, lack of expertise in crucial disciplines, groupthink and bandwagon effects, and selective reporting are some of the causes of these failures. Nevertheless, epidemic forecasting is unlikely to be abandoned. Some (but not all) of these problems can be fixed. Careful modeling of predictive distributions rather than focusing on point estimates, considering multiple dimensions of impact, and continuously reappraising models based on their validated performance may help. If extreme values are considered, extremes should be considered for the consequences of multiple dimensions of impact so as to continuously calibrate predictive insights and decision-making. When major decisions (e.g. draconian lockdowns) are based on forecasts, the harms (in terms of health, economy, and society at large) and the asymmetry of risks need to be approached in a holistic fashion, considering the totality of the evidence.

© 2020 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Initial position

COVID-19 is a major acute crisis with unpredictable consequences. Many scientists have struggled to make forecasts about its impact (Holmdahl & Buckee, 2020). However, despite involving many excellent modelers, best intentions, and highly sophisticated tools, forecasting efforts have largely failed.

Early on, experienced modelers drew parallels between COVID-19 and the Spanish flu (<https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/>). (Accessed 2 June

2020)) that caused >50 million deaths with mean age of death being 28. We all lament the current loss of life. However, as of June 18, the total fatalities are ~450,000 with median age ~80 and typically multiple comorbidities.

Brilliant scientists expected 100,000,000 cases accruing within 4 weeks in the USA (Hains, 2020). Predictions for hospital and ICU bed requirements were also entirely misinforming. Public leaders trusted models (sometimes even black boxes without disclosed methodology) inferring massively overwhelmed health care capacity (Table 1) (IHME COVID-19 health service utilization forecasting team & Murray, 2020). However, very few hospitals were eventually stressed and only for a couple of weeks. Most hospitals maintained largely empty wards, expecting tsunamis that never came. The general population

* Correspondence to: Stanford Prevention Research Center, 1265 Welch Road, Medical School Office Building, Room X306, USA.
 E-mail address: jioannid@stanford.edu (J.P.A. Ioannidis).

Table 1
Some predictions about hospital bed needs and their rebuttal by reality: examples from news coverage of some influential forecasts.

State	Prediction made	What happened
New York (https://www.nytimes.com/2020/04/10/nyregion/new-york-coronavirus-hospitals.html and https://www.forbes.com/sites/sethcohen/2020/05/26/we-all-failed--the-real-reason-behind-ny-governor-andrew-cuomos-surprising-confession/#3e700be06fa5)	“Sophisticated scientists, Mr. Cuomo said, had studied the coming coronavirus outbreak and their projections were alarming . Infections were doubling nearly every three days and the state would soon require an unthinkable expansion of its health care system. To stave off a catastrophe, New York might need up to 140,000 hospital beds and as many as 40,000 intensive care units with ventilators.” 4/10/2020	“But the number of intensive care beds being used declined for the first time in the crisis, to 4,908, according to daily figures released on Friday. And the total number hospitalized with the virus, 18,569, was far lower than the darkest expectations.” 4/10/2020 “Here’s my projection model. Here’s my projection model. They were all wrong. They were all wrong.” Governor Andrew Cuomo 5/25/2020
Tennessee (https://www.nashvillepost.com/business/health-care/article/21127025/covid19-update-hospitalization-projections-drop and https://www.tennessean.com/story/money/industries/health-care/2020/06/04/tennessee-hospitals-expected-lose-3-5-billion-end-june/3139003001/)	“Last Friday, the model suggested Tennessee would see the peak of the pandemic on about April 19 and would need an estimated 15,500 inpatient beds, 2,500 ICU beds and nearly 2,000 ventilators to keep COVID-19 patients alive.”	“Now, it is projecting the peak to come four days earlier and that the state will need 1,232 inpatients beds, 245 ICU beds and 208 ventilators. Those numbers are all well below the state’s current health care capacity.” “Hospitals across the state will lose an estimated \$3.5 billion in revenue by the end of June because of limitations on surgeries and a dramatic decrease in patients during the coronavirus outbreak, according to new estimates from the Tennessee Hospital Association.” 6/4/2020
California (https://www.sacbee.com/news/california/article241621891.html and https://medicalxpress.com/news/2020-04-opinion-hospitals-beds-non-covid-patients.html)	“In California alone, at least 1.2 million people over the age of 18 are projected to need hospitalization from the disease, according to an analysis published March 17 by the Harvard Global Health Institute and the Harvard T.H. Chan School of Public Health... California needs 50,000 additional hospital beds to meet the incoming surge of coronavirus patients, Gov. Gavin Newsom said last week.”	“In our home state of California, for example, COVID-19 patients occupy fewer than two in 10 ICU beds, and the growth in COVID-19-related utilization, thankfully, seems to be flattening out. California’s picture is even sunnier when it comes to general hospital beds. Well under five percent are occupied by COVID-19 patients.”

was locked and placed in horror-alert to save health systems from collapsing. Tragically, many health systems faced major adverse consequences, not by COVID-19 cases overload, but for very different reasons. Patients with heart attacks avoided hospitals for care (De Filippo, D’Ascenzo, Angelini, et al., 2020), important treatments (e.g. for cancer) were unjustifiably delayed (Sud et al., 2020) and mental health suffered (Moser, Glaus, Frangou, et al., 2020). With damaged operations, many hospitals started losing personnel, reducing their capacity to face future crises (e.g. a second wave). With massive new unemployment, more people may lose health insurance. The prospects of starvation and of lack of control of other infectious diseases (such as tuberculosis, malaria, and childhood communicable diseases where vaccination is hindered by COVID-19 measures) are dire (Ioannidis, 2020; Melnick & Ioannidis, 2020).

Modeling resurgence after reopening also failed (Table 2). For example, a Massachusetts General Hospital model (<https://www.massgeneral.org/news/coronavirus/covid-19-simulator>. (Accessed 2 June 2020)) predicted over 23,000 deaths within a month of Georgia reopening – the actual deaths were 896.

Table 3 lists some main reasons underlying this forecasting failure. Unsurprisingly, models failed when they used more speculation and theoretical assumptions and tried to predict long-term outcomes; for example, using early SIR-based models to predict what would happen in the entire season. However, even forecasting built directly on data alone fared badly (Chin et al., 2020b; Marchant, Samia, Rosen, Tanner, & Cripps, 2020), failing not only in ICU bed predictions (Fig. 1) but also in next day death predictions when issues of long-term chaotic behavior

Table 2
Forecasting what will happen after reopening.

PREDICTION FOR REOPENING	WHAT ACTUALLY HAPPENED
<p>“Results indicate that lifting restrictions too soon can result in a second wave of infections and deaths. Georgia is planning to open some businesses on April 27th. The tool shows that COVID-19 is not yet contained in Georgia and even lifting restrictions gradually over the next month can result in over 23,000 deaths.”</p> <p>Massachusetts General Hospital News, April 24, 2020</p> <p>(https://www.massgeneral.org/news/coronavirus/COVID-19-simulator)</p>	<p>Number of deaths over one month: 896 instead of the predicted 23,000</p>
<p>“administration is privately projecting a steady rise in the number of coronavirus cases and deaths over the next several weeks. The daily death toll will reach about 3,000 on June 1, according to an internal document obtained by The New York Times, a 70 percent increase from the current number of about 1,750.</p> <p>The projections, based on government modeling pulled together by the Federal Emergency Management Agency, forecast about 200,000 new cases each day by the end of the month, up from about 25,000 cases a day currently.”</p> <p>New York Times, May 4, 2020 (https://www.nytimes.com/2020/05/04/us/coronavirus-live-updates.html)</p>	<p>Number of daily deaths on June 1: 731 instead of the predicted 3,000, i.e. a 60% decrease instead of 70% increase</p> <p>Number of daily new cases on May 31: 20,724 instead of the predicted 125,000, i.e. a 15% decrease instead of 700% increase</p>
<p>“According to the Penn Wharton Budget Model (PWBM), reopening states will result in an additional 233,000 deaths from the virus — even if states don’t reopen at all and with social distancing rules in place. This means that if the states were to reopen, 350,000 people in total would die from coronavirus by the end of June, the study found.”</p> <p>Yahoo, May 3, 2020 (https://www.yahoo.com/now/reopening-states-will-cause-233000-more-people-to-die-from-coronavirus-according-to-wharton-model-120049573.html)</p>	<p>Based on JHU dashboard death count, number of additional deaths as of June 30 was 5,700 instead of 233,000, i.e. total deaths was 122,700 instead of 350,000. It is unclear also whether any of the 5,700 deaths were due to reopening rather than error in the original model calibration of the number of deaths without reopening.</p>
<p>“Dr. Ashish Jha, the director of the Harvard Global Health Institute, told CNN’s Wolf Blitzer that the current data shows that somewhere between 800 to 1000 Americans are dying from the virus daily, and even if that does not increase, the US is poised to cross 200,000 deaths sometime in September.</p> <p>“I think that is catastrophic. I think that is not something we have to be fated to live with,” Jha told CNN. “We can change the course. We can change course today.”</p> <p>“We’re really the only major country in the world that opened back up without really getting our cases as down low as we really needed to,” Jha told CNN.”</p> <p>Business Insider, June 10, 2020 (https://www.businessinsider.com/harvard-expert-predicts-coronavirus-deaths-in-us-by-september-2020-6)</p>	<p>Within less than 4 weeks of this quote, the number of daily deaths was much less than the 800–1000 quote (516 daily average for the week ending July 4). Then it increased again to over 1000 daily average in the first three weeks in August and then it decreased again to 710 daily average by the last week of September. Predictions are precarious with such volatile behavior of the epidemic wave.</p>

do not come into play (Figs. 2 and 3). Even for short-term forecasting when the epidemic wave waned, models presented confusingly diverse predictions with huge uncertainty (Fig. 4).

Failure in epidemic forecasting is an old problem. In fact, it is surprising that epidemic forecasting has retained much credibility among decision-makers, given its dubious track record. Modeling for swine flu predicted 3,100–65,000 deaths in the UK (<https://www.theguardian.com>

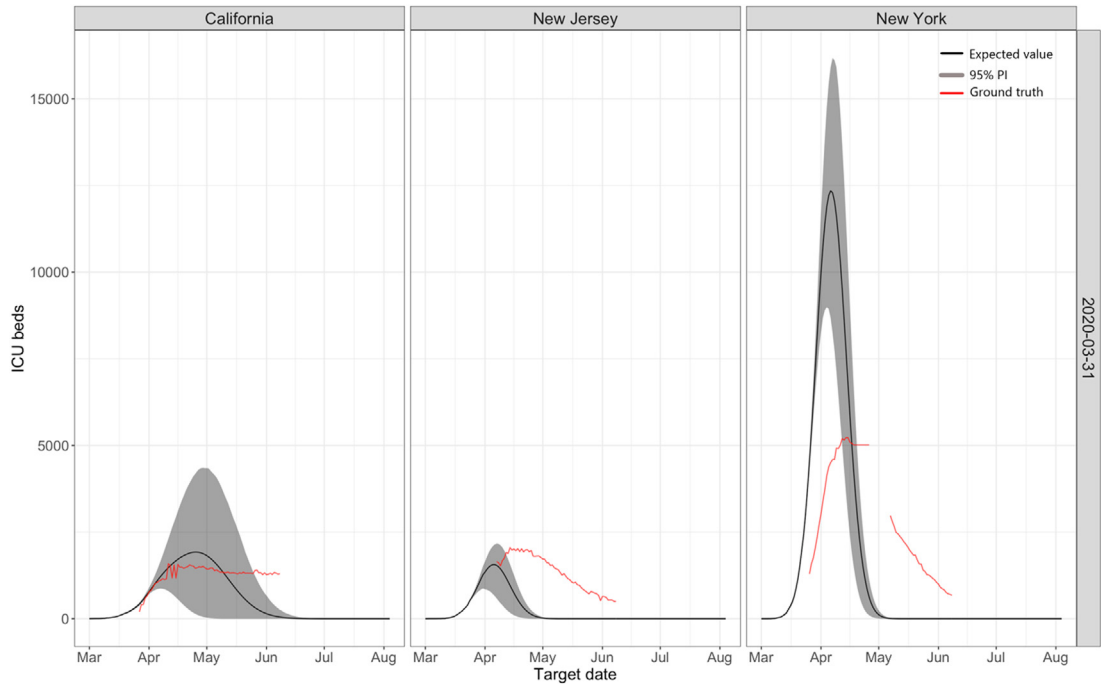


Fig. 1. Predictions for ICU beds made by the IHME models on March 31 for three states: California, New Jersey, and New York. For New York, the model initially over predicted enormously, and then it under predicted. For New Jersey, a neighboring state, the model started well but then it underpredicts, while for California it predicted a peak which never eventuated.

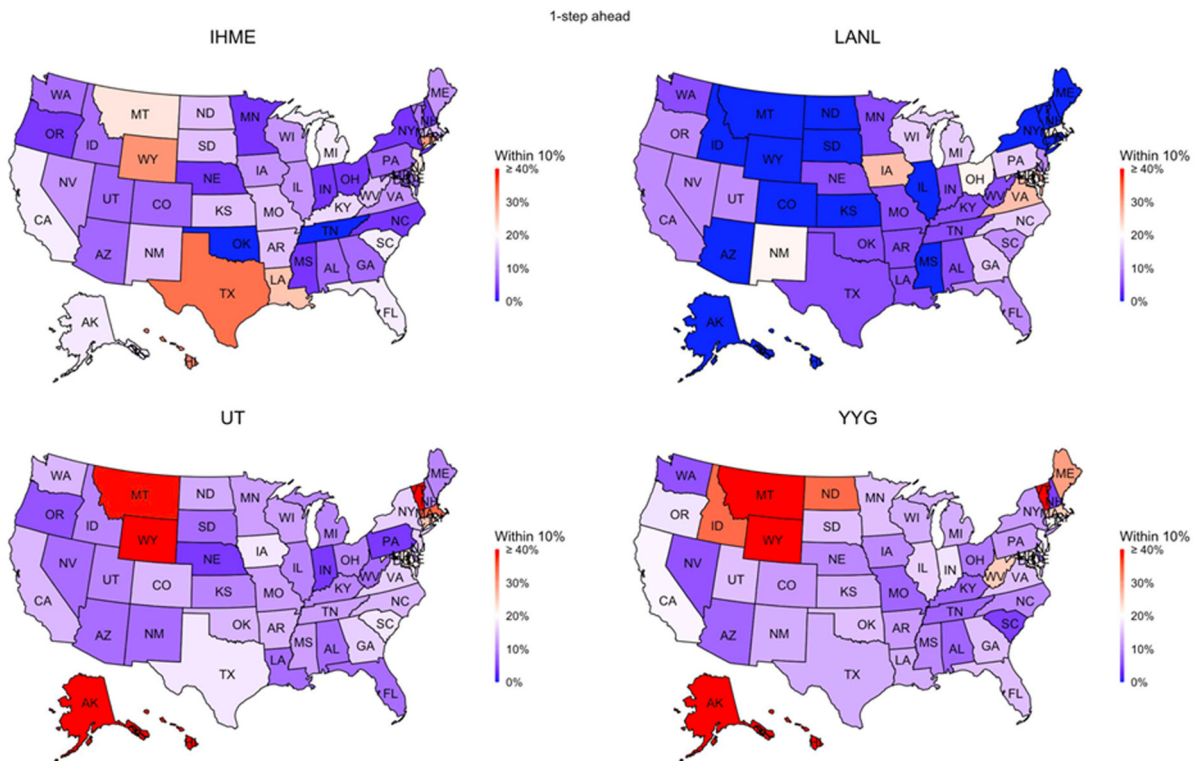


Fig. 2. Performance of four data-driven models, IHME, YYG, UT, and LANL, used to predict COVID-19 death counts by state in the USA for the following day. That is, these were predictions made only 24 h in advance of the day in question. The Figure shows the percentage of times that a particular model's prediction was within 10% of the ground truth by state. All models failed in terms of accuracy; for the majority of states, this figure was less than 20%.

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 sites. 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.

(continued on next page)

[/uk/2009/jul/16/swine-flu-cases-rise-britain](#). (Accessed 2 June 2020)). Eventually, 457 deaths occurred (UK government, 2009). Models on foot-and-mouth disease by top scientists in top journals (Ferguson, Donnelly, & Anderson, 2001a, 2001b) were subsequently questioned (Kitching,

Thrusfield, & Taylor, 2006) by other scientists challenging why up to 10 million animals had to be slaughtered. Predictions for bovine spongiform encephalopathy expected up to 150,000 deaths in the UK (Ferguson, Ghani, Donnelly, Hagensaaers, & Anderson, 2002). However, the

Table 3 (continued).

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 be 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

lower bound predicted as low as 50 deaths (Ferguson et al., 2002), which is a figure close to eventual fatalities. Predictions may work in “ideal”, isolated communities with homogeneous populations, not the complex current global world.

Despite these obvious failures, epidemic forecasting continued to thrive, perhaps because vastly erroneous predictions typically lacked serious consequences. In fact, erroneous predictions may have even been useful. A wrong, doomsday prediction may incentivize people towards better personal hygiene. Problems emerge when public leaders take (wrong) predictions too seriously, considering them crystal balls without understanding their

uncertainty and the assumptions made. Slaughtering millions of animals may aggravate animal business stakeholders – but most citizens are not directly affected. However, with COVID-19, espoused wrong predictions can devastate billions of people in terms of the economy, health, and societal turmoil at-large.

Let us be clear: even if millions of deaths did not happen this season, they may happen with the next wave, next season, or some new virus in the future. A doomsday forecast may come in handy to protect civilization when and if calamity hits. However, even then, we have little evidence that aggressive measures focusing only on a few dimensions of impact actually reduce death toll

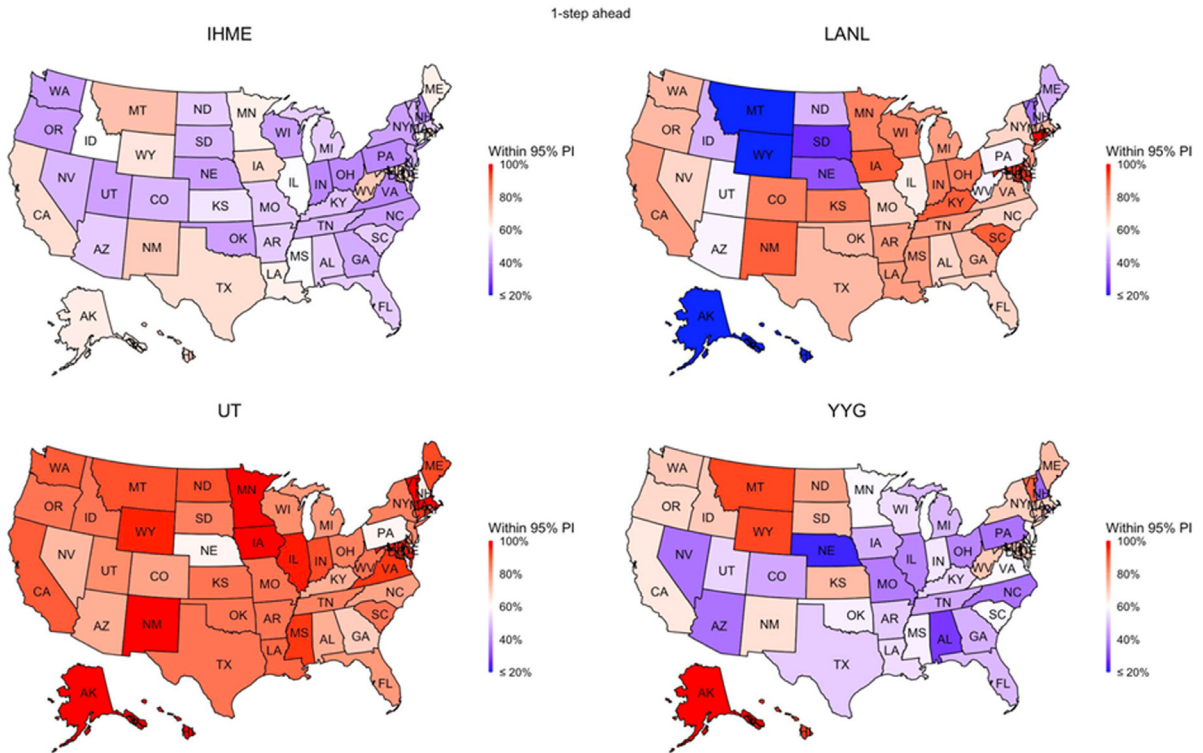


Fig. 3. Performance of the same models examined in Fig. 2 in terms of their uncertainty quantification. If a model assessment of uncertainty is accurate, then we would expect 95% of the ground truth values to fall within the 95% prediction interval. Only one of the 4 models (the UT model) approached this level of accuracy.

and do more good than harm. We need models which incorporate multicriteria objective functions. Isolating infectious impact, from all other health, economic, and social impacts is dangerously narrow-minded. More importantly, with epidemics becoming easier to detect, opportunities for declaring global emergencies will escalate. Erroneous models can become powerful, recurrent disruptors of life on this planet. Civilization is threatened by epidemic incidentalomas.

Cirillo and Taleb thoughtfully argue (Cirillo & Taleb, 2020) that when it comes to contagious risk, we should take doomsday predictions seriously: major epidemics follow a fat-tail pattern and extreme value theory becomes relevant. Examining 72 major epidemics recorded through history, they demonstrate a fat-tailed mortality impact. However, they analyze only the 72 most-noticed outbreaks, which is a sample with astounding selection bias. For example, according to their dataset, the first epidemic originating from sub-Saharan Africa did not occur until 1920 AD, namely HIV/AIDS. The most famous outbreaks in human history are preferentially selected from the extreme tail of the distribution of all outbreaks. Tens of millions of outbreaks with a couple deaths must have happened throughout time. Around hundreds of thousands might have claimed dozens of fatalities. Thousands of outbreaks might have exceeded 1000 fatalities. Most eluded the historical record. The four garden variety coronaviruses may be causing such outbreaks every year (Patrick, Petric, Skowronski, et al., 2006; Walsh, Shin, &

Falsey, 2013). One of them, OC43 seems to have been introduced in humans as recently as 1890, probably causing a “bad influenza year” with over a million deaths (Vijgen et al., 2005). Based on what we know now, SARS-CoV-2 may be closer to OC43 than SARS-CoV-1. This does not mean it is not serious: its initial human introduction can be highly lethal, unless we protect those at risk.

A heavy tail distribution ceases to be as heavy as Taleb imagines when the middle of the distribution becomes much larger. One may also argue that pandemics, as opposed to epidemics without worldwide distribution, are more likely to be heavy-tailed. However, the vast majority of the 72 contagious events listed by Taleb were not pandemics, but localized epidemics with circumscribed geographic activity. Overall, when a new epidemic is detected, it is even difficult to pinpoint which distribution of which known events it should be mapped against.

Blindly acting based on extreme value theory alone would be sensible if we lived in the times of the Antonine plague or even in 1890, with no science to identify the pathogen, elucidate its true prevalence, estimate accurately its lethality, and carry out good epidemiology to identify which people and settings are at risk. Until we accrue this information, immediate better-safe-than-sorry responses are legitimate, trusting extreme forecasts as possible (not necessarily likely) scenarios. However, caveats of these forecasts should not be ignored (Holmdahl & Buckee, 2020; Jewell, Lewnard, & Jewell, 2020) and new evidence on the ground truth needs continuous

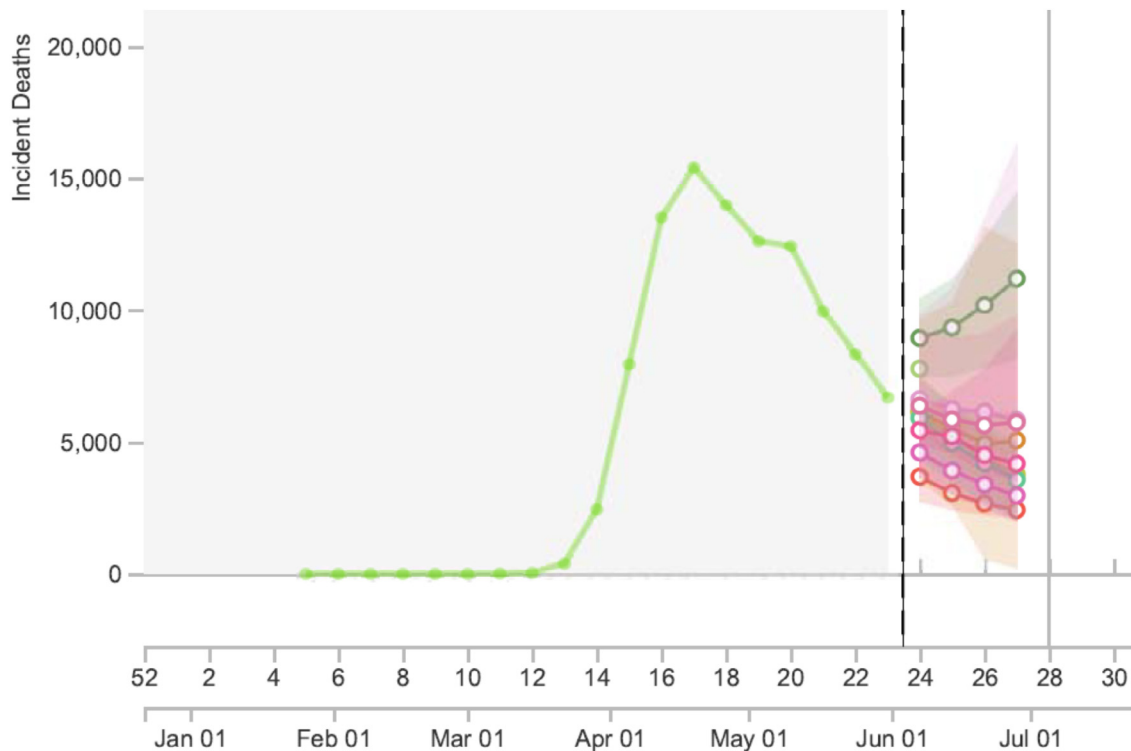


Fig. 4. Snapshot from <https://reichlab.io/covid19-forecast-hub/> (a very useful site that collates information and predictions from multiple forecasting models) as of 11.14 AM PT on June 3, 2020. Predictions for number of US deaths during week 27 (only ~3 weeks downstream) with these 8 models ranged from 2419 to 11190, which is a 4.5-fold difference, and the spectrum of 95% confidence intervals ranged from fewer than 100 deaths to over 16,000 deaths, which is almost a 200-fold difference.

reassessment. Upon acquiring solid evidence about the epidemiological features of new outbreaks, implausible, exaggerated forecasts (Ioannidis, 2020d) should be abandoned. Otherwise, they may cause more harm than the virus itself.

2. Further thoughts – analogies, decisions of action, and maxima

The insightful recent essay of Taleb (2020) offers additional opportunities for fruitful discussion.

2.1. Point estimate predictions and technical points

Taleb (2020) ruminates on the point of making point predictions. Serious modelers (whether frequentist or Bayesian) would never rely on point estimates to summarize skewed distributions. Even an early popular presentation (Huff & Geis, 1954) from 1954 has a figure (see page 33) with striking resemblance to Taleb's Fig. 1 (Taleb, 2020). In a Bayesian framework, we rely on the full posterior predictive distribution, not single points (Tanner, 1996). Moreover, Taleb's choice of a three-parameter Pareto distribution is peculiar. It is unclear whether this model provides a measurably better fit to his (hopelessly biased) pandemic data (Cirillo & Taleb, 2020) than, say, a two parameter Gamma distribution fitted to log counts. Regardless, either skewed distribution would then have to be modified to allow for the use of all available sources

of information in a logically consistent fully probabilistic model, such as via a Bayesian hierarchical model (which can certainly be formulated to accommodate fat tails if needed). In this regard, we note that upon examining the NY daily death count data studied in Chin et al. (2020b), these data are found to be characterized as stochastic rather than chaotic (Toker, Sommer, & D'Esposito, 2000). Taleb seems to fit an unorthodox model, and then abandons all effort to predict anything. He simply assumes doomsday has come, much like a panic-driven Roman would have done in the Antonine plague, lacking statistical, biological, and epidemiological insights.

2.2. Should we wait for the best evidence before acting?

Taleb (2020) caricatures the position of a hotly debated mid-March op-ed by one of us, Ioannidis (2020a) suggesting that it "made statements to the effect that one should wait for "more evidence" before acting with respect to the pandemic", which is an obvious distortion of the op-ed. Anyone who reads the op-ed unbiasedly realizes that it says exactly the opposite. It starts with the clear, unquestionable premise that the pandemic is taking hold and is a serious threat. Immediate lockdown certainly makes sense when an estimated 50 million deaths are possible. This is stated emphatically on multiple occasions these days in interviews in multiple languages -for examples see Ioannidis, 2020b, 2020c). Certainly, adverse consequences of short-term lockdown cannot match 50

million lives. However, better data can help recalibrate estimates, re-assessing downstream the relative balance of benefits and harms of longer-term prolongation of lockdown. That re-appraised balance changed markedly over time (Ioannidis, 2020).

Another gross distortion propagated in social media is that the op-ed (Ioannidis, 2020a) had supposedly predicted that only 10,000 deaths will happen in the USA as a result of the pandemic. The key message of the op-ed was that we lack reliable data, that is, we do not know. The self-contradicting misinterpretation as “we don’t know, but actually we do know that 10,000 deaths will happen” is impossible. The op-ed discussed two extreme scenarios to highlight the tremendous uncertainty absent reliable data: an overtly optimistic scenario of only 10,000 deaths in the US and an overtly pessimistic scenario of 40,000,000 deaths. We needed reliable data, quickly, to narrow this vast uncertainty. We did get data and did narrow uncertainty. Science did work eventually, even if forecasts, including those made by one of us (confessed and discussed in Appendix), failed.

2.3. Improper and proper analogies of benefit-risk

Taleb (2020) offers several analogies to assert that all precautionary actions are justified in pandemics, deriding “waiting for the accident before putting the seat belt on, or evidence of fire before buying insurance” (Taleb, 2020). The analogies assume that the cost of precautionary actions are small in comparison to the cost of the pandemic, and that the consequences of the action have little impact on it. However, precautionary actions can backfire severely when they are misinformed. In March, modelers were forecasting collapsed health systems; for example 140,000 beds would be needed in New York, when only a small fraction were available. Precautionary actions damaged the health system, increased COVID-19 deaths (AP counts: over 4500 virus patients sent to NY nursing homes, 2020), and exacerbated other health problems (Table 4).

Seat belts cost next to nothing to produce in cars and have unquestionable benefits. Despite some risk compensation and some excess injury with improper use, seat belts eventually prevent ~50% of serious injuries and deaths (National Highway Traffic Safety Administration, 2017). Measures for pandemic prevention equivalent to seat belts in terms of benefit-harm profile are simple interventions like hand washing, respiratory etiquette, and mask use in appropriate settings: large proven benefit, no/little harm/cost (Jefferson, Del Mar, Dooley, et al., 2011; Saunders-Hastings, Crispo, Sikora, & Krewski, 2017). Even before the COVID-19 pandemic, we had randomized trials showing 38% reduced odds of influenza infection with hand washing and (non-statistically significant, but possible) 47% reduced odds with proper mask wearing (Saunders-Hastings et al., 2017). Despite lack of trials, it is sensible and minimally disruptive to avoid mass gatherings and decrease unnecessary travel. Prolonged draconian lockdown is not equivalent to seat belts. It resembles forbidding all commute.

Similarly, fire insurance offers a misleading analogy. Fire insurance makes sense only at reasonable price. Draconian prolonged lockdown may be equivalent to paying fire insurance at a price higher than the value of the house.

2.4. Mean, observed maximum, and more than the observed maximum

Taleb refers to the Netherlands where maximum values for flooding, not the mean, are considered (Taleb, 2020). Anti-flooding engineering has substantial cost but a favorable decision-analysis profile after considering multiple types of impact. Lockdown measures were decided based on examining only one type of impact, COVID-19. Moreover, the observed flooding maximum to-date does not preclude even higher future values. Netherlands aims to avoid devastation from floods occurring once every 10,000 years in densely populated areas (https://en.wikipedia.org/wiki/Flood_control_in_the_Netherlands. (Accessed 18 June 2020)). A more serious flooding event (e.g. one that occurs every 20,000 years) may still submerge the Netherlands next week. However, prolonged total lockdown is not equivalent to building higher sea walls. It is more like abandoning the country - asking the Dutch to immigrate, because their land is quite unsafe.

Other natural phenomena also exist where high maximum risks are difficult to pinpoint and where new maxima may be reached. For example, following Taleb’s argumentation, one should forbid living near active volcanoes. Living at the Santorini caldera is not exciting, but foolish: that dreadful island should be summarily evacuated. The same applies to California: earthquake devastation may strike any moment. Prolonged lockdown zealots might barely accept a compromise: whenever substantial seismic activity occurs, California should be temporarily evacuated until all seismic activity ceases.

Furthermore, fat-tailed uncertainty and approaches based on extreme value theory may be useful before a potentially high-risk phenomenon starts and during its early stages. However, as more data accumulate and the high-risk phenomenon can be understood more precisely with plenty of data, the laws of large numbers may apply and stochastic rather than chaotic approaches may become more relevant and useful than continuing to assume unlikely extremes. Further responses to Taleb (2020) appear in Table 5.

3. Moving forward and learning from the COVID-19 pandemic and from our mistakes

3.1. How do we move forward to deal with the COVID-19 threat?

The short answer is ‘using science and more reliable data’. We can choose measures with favorable benefit-risk ratio, when we consider together multiple types of impact, not only on COVID-19, but on health as a whole, as well as society and economy.

Currently, we know that approximately half of the COVID-19 deaths in Europe and the USA affected nursing

Table 4

Adverse consequences of precautionary actions, expecting excess of hospitalization and ICU needs (as forecasted by multiple models).

PRECAUTIONARY ACTION	JUSTIFICATION	WHAT WENT WRONG
Stop elective procedures, delay other treatments	Focus available resources on preparing for the COVID-19 onslaught	Treatment for major conditions like cancer were delayed, (Sud et al., 2020) effective screening programs were cancelled, procedures not done on time had suboptimal outcomes
Send COVID-19 patients to nursing homes	Acute hospital beds are needed for the predicted COVID-19 onslaught, models predict hospital beds will not be enough	Thousands of COVID-19 infected patients were sent to nursing homes (AP counts: over 4500 virus patients sent to NY nursing homes, 2020) where large numbers of ultra-vulnerable individuals are clustered together; may have massively contributed to eventual death toll
Inform the public that we are doing our best, but it is likely that hospitals will be overwhelmed by COVID-19	Honest communication with the general public	Patients with major problems like heart attacks did not come to the hospital to be treated, (De Filippo et al., 2020) while these are diseases that are effectively treatable only in the hospital; an unknown, but probably large share of excess deaths in the COVID-19 weeks were due to these causes rather than COVID-19 itself (Grisin, 2020; Mansour et al., 2020)
Re-orient all hospital operations to focus on COVID-19	Be prepared for the COVID-19 wave, strengthen the response to crisis	Most hospitals saw no major COVID-19 wave and also saw a massive reduction in overall operations with major financial cost, leading to furloughs and lay-off of personnel; this makes hospitals less prepared for any major crisis in the future

home residents (Danis, Fonteneau, Georges, et al., 2020; Nursing Homes & Assisted Living Facilities Account for 45% of COVID-19 Deaths). Another sizeable proportion were nosocomial infections (Boccia, Ricciardi, & Ioannidis, 2020). If we protect these locations with draconian hygiene measures and intensive testing, we may avert 70% of the fatalities without large-scale societal disruption and without adverse consequences on health. Other high-risk settings, for example, prisons, homeless shelters, and meat-processing plants, also need aggressive protection. For the rest of the population, we have strong evidence of a very steep age gradient with ~1000-fold differences in death risk for people >80 years of age versus children (Ioannidis, Axford, & Contopoulos-Ioannidis, 2020). We also have also detailed insights on how different background diseases modify COVID-19 risk for death or other serious outcomes (Williamson et al., 2020). We can use hygiene and some least disruptive distancing measures to protect people. We can use intensive testing (i.e. again, use science) to detect resurgence of epidemic activity and extinguish it early – the countries that most successfully faced the first wave, namely Singapore and Taiwan, did exactly that highly successfully. We can use data to track how the epidemic and its impact evolve. Data can help inform more granular models and titrate decisions considering distributions of risk (Fig. 5) (Williamson et al., 2020).

3.2. Abandon or improve epidemic forecasting?

Poorly performing models and models that perform well for only one dimension of impact can cause harm. It is not just an issue of academic debate, it is an issue of potentially devastating, wrong decisions (Jefferson et al., 2011). Taleb (2020) seems self-contradicting: does he espouse abandoning all models (as they are so wrong) or using models but always assuming the worst? However, there is no single worst scenario, but a centile of the distribution: should we prepare for an event that has 0.1%, 0.001%, or 0.000000000001% chance of happening? Paying what price in harmful effects?

Abandoning all epidemic modeling appears too unrealistic. Besides identifying the problems of epidemic modeling, Table 3 also offers suggestions on addressing some of them.

To summarize, here are some necessary (although not always sufficient) targets for amendments:

- Invest more on collecting, cleaning, and curating real, unbiased data, and not just theoretical speculations
- Model the entire predictive distribution, with particular focus on accurately quantifying uncertainty
- Continuously monitor the performance of any model against real data and either re-adjust or discard models based on accruing evidence.

Table 5

Taleb's main statements and our responses.

<p>Forecasting single variables in fat tailed domains is in violation of both common sense and probability theory.</p>	<p><i>Serious</i> statistical modelers (whether frequentist or Bayesian) would never rely on point estimates to summarize a skewed distribution. Using data as part of a decision process is not a violation of common sense, irrespective of the distribution of the random variable. Possibly using only data and ignoring what is previously known (or expert opinion or physical models) may be unwise in small data problems. We advocate a Bayesian approach, incorporating different sources of information into a logically consistent fully probabilistic model. We agree that higher order moments (or even the first moment in the case of the Cauchy distribution) do not exist for certain distributions. This does not preclude making probabilistic statements such as $P(a < X < b)$.</p>
<p>Pandemics are extremely fat tailed.</p>	<p>Yes, and so are many other phenomena. The distribution of financial rewards from mining activity, for example, is incredibly fat tailed and very asymmetric. As such, it is important to accurately quantify the entire distribution of forecasts. From a Bayesian perspective, we can rely on the posterior distribution (as well as the posterior predictive distribution) as the basis of statistical inference and prediction (Tanner, 1996).</p>
<p>Science is not about making single points predictions but understanding properties (which can <i>sometimes</i> be tested by single points).</p>	<p>We agree and that is why the focus should be on the entire predictive distribution, and why we should be flexible in the way in which we model and estimate this distribution. Bayesian hierarchical models (which can be formulated to account for fat tails, if need be) may allow using all available sources of information in a logically consistent fully probabilistic model.</p>
<p>Risk management is concerned with tail properties and distribution of extrema, not averages or survival functions.</p>	<p>Quality data and calibrated (Bayesian) statistical models may be useful in estimating the behaviour of a random variable across the whole spectrum of outcomes, not just point estimates of summary statistics. While the three parameter Pareto distribution can be developed based on interesting mathematics, it is not clear that it will provide a measurably better fit to skewed data (e.g. pandemic data in ref. 19) than would a two parameter Gamma distribution fitted to the log counts. It is certainly not immediately obvious how to generalize either skewed distribution to allow for the use of all available sources of information in a logically consistent fully probabilistic model. In this regard, we note that upon examining the NY daily death count data studied in (Chin et al., 2020b), these data are found to be characterized as stochastic rather than chaotic. (Toker et al., 2000)</p>
<p>Naive fortune cookie evidentiary methods fail to work under both risk management and fat tails as absence of evidence can play a large role in the properties.</p>	<p>A passenger may well get off the plane if it is on the ground and the skills of the pilot are in doubt, but what if he awakes to find he is on a nonstop from JFK to LAX? The poor passenger can stay put, cross her/his fingers, say a few prayers, or can get a parachute and jump; assuming s/he is able and willing to open the exit door in midflight. The choice is not so easy when there are considerable risks associated with either decision that need to be made in real time. We argue that acquiring further information on the pilot's skill level, perhaps from the flight attendant as s/he strolls down the aisle with the tea trolley, as well as checking that the parachute under the seat (if available) has no holes, would be prudent courses of action. This is exactly the situation that faced New York- they did not arrange to be ground zero of the COVID-19 pandemic in the US. Various models forecast very high demand for ICU beds in New York state. As a result of this forecast, a decision was made to send COVID-19 patients to nursing homes, with tragic consequences.</p>

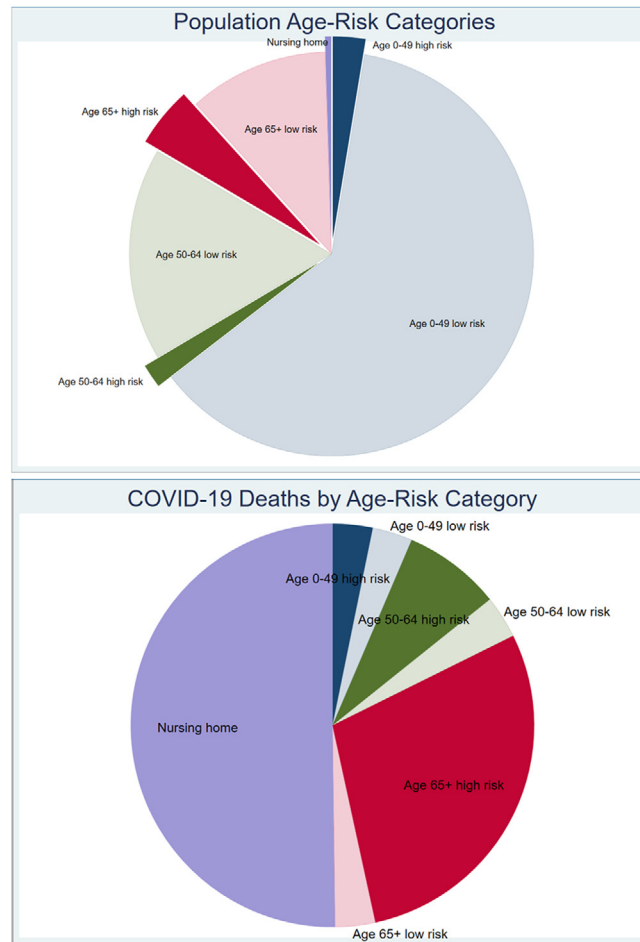


Fig. 5. Population age-risk categories and COVID-19 deaths per age-risk category. The illustration uses estimates for a symptomatic case fatality rate of 0.05% in ages 0–49, 0.2% in ages 50–64, and 1.3% in ages 65 and over, similar to the CDC main planning scenario (<https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>). It also assumes that 50% of infections are asymptomatic in ages 0–49, 30% are asymptomatic in ages 50–64, and 10% are asymptomatic in ages 65 and over. Furthermore, it assumes that among people in nursing homes and related facilities (~0.5% of the population in the USA), the infection fatality rate is 26%, as per Arons, Hatfield, Reddy, Kimball, James, et al. (2020). Finally, it assumes that some modest prognostic model is available where 4% of highest-risk people 0–49 years old explain 50% of the death risk in that category, the top 10% explains 70% of the deaths in the 50–64 years category, and the top 30% explains 90% of the risk in the 65 and above category. Based on available prognostic models (e.g. Williamson et al. (2020)), this prognostic classification should be readily attainable. As shown, <10% of the population is at high risk (shown with dense-colors and thus worth special protection and more aggressive measures), and these people account for >90% of the potential deaths. More than 90% of the population could possibly continue with non-disruptive measures as they account for only <10% of the total potential deaths.

- Incorporate the best epidemiological estimates on age structure and comorbidities in the modeling
- Focus on quality-adjusted life-years rather than deaths
- Avoid unrealistic assumptions about the benefits of interventions; do not hide model failure behind implausible intervention effects
- Enhance transparency about the methods
- Share code and data
- Use up-to-date and well-vetted tools and processes that minimize the potential for error through auditing loops in the software and code
- Promote interdisciplinarity and ensure that the modelers’ teams are diversified and solidly grounded in terms of subject matter expertise
- Maintain an open-minded approach and acknowledge that most forecasting is exploratory, subjective, and non-pre-registered research
- Beware of unavoidable selective reporting bias

Of interest, another group of researchers have reached (Saltelli et al., 2020) almost identical conclusions, and their recommendations are largely overlapping. Hype in forecasting can be dangerous.

Importantly, not all problems can be fixed. At best, epidemic models offer only tentative evidence. Great caution and nuance are still needed. Models that use reliable data, that are validated and continuously reappraised for their performance in real-time, and that combine multiple dimensions of impact may have more utility. A good starting point is to acknowledge that problems exist. Serious

scientists who have published poorly performing models should acknowledge this. They may also correct or even retract their papers, receiving credit and congratulations, not blame, for corrections/retractions. The worst nightmare would be if scientists and journals insist that prolonged draconian measures cause the massive difference between predictions and eventual outcomes. Serious scientists and serious journals (Flaxman et al., 2020) are unfortunately flirting with this slippery, defensive path (Chin, Ioannidis, Tanner, & Cripps, 2020a; Homburg & Kuhbandner, 2020). Total lockdown is a bundle of dozens of measures. Some may be very beneficial, but some others may be harmful. Hiding uncertainty can cause major harm downstream and leaves us unprepared for the future. For papers that fuel policy decisions with major consequences, transparent availability of data, code, and named peer-review comments is also a minimum requirement.

Calibrating model predictions for looking at extremes rather than just means is sensible, especially in early days of pandemics, when much is unknown about the virus and its epidemiological footprint. However, when calibration/communication on extremes is adopted, one should also consider similar calibration for the potential harms of adopted measures. For example, tuberculosis has killed 1 billion people in the last 200 years, it still kills 1.5 million people (mostly young and middle age ones) annually, and prolonged lockdown may cause 1.4 million extra tuberculosis deaths between 2020–2025 (<https://www.healio.com/news/infectious-disease/20200506/covid19-will-set-fight-against-tb-back-at-least-5-years>. (Accessed 20 June 2020)). Measles has killed about 200 million people in the last 150 years; disrupted MMR vaccination may fuel lethal recrudescence (<https://apps.who.int/iris/handle/10665/331590>. (Accessed 20 June 2020)). Extreme case predictions for COVID-19 deaths should be co-examined with extreme case predictions for deaths and impacts from lockdown-induced harms. Models should provide the big picture, covering multiple dimensions. Similar to COVID-19, as more reliable data accrue, predictions on these other dimensions should also be corrected accordingly.

Eventually, it is probably impossible (and even undesirable) to ostracize epidemic forecasting, despite its failures. Arguing that forecasting for COVID-19 has failed should not be misconstrued to mean that science has failed. Developing models in real time for a novel virus, with poor quality data, is a formidable task, and the groups who attempted this and made public their predictions and data in a transparent manner should be commended. We readily admit that it is far easier to criticize a model than to build one. It would be horrifically retrograde if this debate ushers in a return to an era where predictions, on which huge decisions are made, are kept under lock and key (e.g. by the government - as is the case in Australia).

3.3. Learning from the COVID-19 pandemic and from our mistakes

We wish to end on a more positive note, namely where we feel forecasting has been helpful. Perhaps the

biggest contribution of these models is that they serve as a springboard for discussions and debates. Dissecting the variation in the performances of various models (e.g. casting a sharp eye to circumstances where a particular model excelled) can be highly informative, and a systematic approach to the development and evaluation of such models is needed (Chin et al., 2020b). This demands a coherent approach to collecting, cleaning, and curating data, as well as a transparent approach to evaluating the suitability of models with regard to predictions and forecast uncertainty.

What we have learned from the COVID-19 pandemic can be passed to future generations that hopefully should be better prepared to deal with a new, different pandemic, learning from our failures. There is no doubt that, an explosive literature of models and forecasting will emerge again as soon as a new pandemic is suspected. However, we can learn from our current mistakes to be more cautious with interpreting, using, and optimizing these models. Being more cautious does not mean acting indecisively, but it requires looking at the totality of the data; considering multiple types of impact; involving scientists from very different disciplines; replacing speculations, theories and assumptions with real, empirical data as quickly as possible; and modifying and aligning decisions to the evolving best evidence.

In the current pandemic, we largely failed to protect people and settings at risk. We could have done much better in this regard. It is difficult to correct mistakes that have already led to people dying, but we can avoid making the same mistakes in future pandemics from different pathogens. We can avoid making the same mistakes even for COVID-19 going forward, as this specific pandemic has not ended as we write. In fact, its exact eventual impact is still unknown. For example, the leader of the US task force, Dr. Anthony Fauci, recently warned of reaching 100,000 COVID-19 US cases per day ([US could see 100, 000 new COVID-19 cases per day, Fauci says, 2020](#)). Maybe this prediction is already an underestimate, because with over 50,000 cases diagnosed per day in early July 2020, the true number of infections may be many times larger. There is currently wide agreement that the number of infections in many parts of the United States is more than 10 times higher than the reported rates ([Actual Coronavirus Infections Vastly Undercounted, 2020](#)). We do have interventions that can prevent or reduce the resurgence of the epidemic wave. Moreover, we know that 100,000 cases in healthy children and young adults may translate to almost 0 deaths. Conversely, 100,000 cases in high-risk susceptible individuals and settings may translate to many thousands of deaths. We can use science to extinguish epidemic waves in many circumstances. If extinguishing these waves is not possible, we could at least have them spend their flare on settings where they carry minimal risk. The same forecast for the number of cases may vary 1000-fold or more in terms of outcomes that matter. We should use forecasting, along with many other tools and various types of evidence to improve outcomes that matter.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank Vincent Chin for his helpful discussions and for providing Figs. 1–3.

Appendix. Box 1. John Ioannidis: a fool's confession and dissection of a forecasting failure

"If I were to make an informed estimate based on the limited testing data we have, I would say that COVID-19 will result in fewer than 40,000 deaths this season in the USA" - my quote appeared on April 9 in CNN and Washington Post based on a discussion with Fareed Zakaria a few days earlier. Fareed is an amazingly charismatic person and our discussion covered a broad space. While we had focused more on the need for better data, when he sent me the quote that he planned to use, I sadly behaved like an expert and endorsed it. Journalists and the public want certainty, even when there is no certainty.

Here is an effort to dissect why I was so wrong. Behaving like an expert (i.e. a fool) was clearly the main reason. But there were additional contributing reasons. When I made that tentative quote, I had not considered the impact of the new case definition of COVID-19 and of COVID-19 becoming a notifiable disease (https://cdn.ymaws.com/www.cste.org/resource/resmgr/2020ps/Interim-20-ID-01_COVID-19.pdf. (Accessed 20 June 2020)), despite being aware of the Italian experience (Boccia et al., 2020) where almost all counted "COVID-19 deaths" also had other concomitant causes of death/comorbidities. "COVID-19 death" now includes not only "deaths by COVID-19" and "deaths with COVID-19", but even deaths "without COVID-19 documented". Moreover, I had not taken seriously into account weekend reporting delays in death counts. Worse, COVID-19 had already started devastating nursing homes in the USA by then, but the nursing home data were mostly unavailable. I could not imagine that despite the Italian and Washington state (Roxby et al., 2020) experience, nursing homes were still unprotected. Had I known that nursing homes were even having COVID-19 patients massively transferred to them, I would have escalated my foolish quote several fold.

There is more to this: since mid-March, I wrote an article alerting that there are two settings where the new virus can be devastating and that we need to protect, using draconian measures, nursing homes and hospitals. Over several weeks, I tried unsuccessfully to publish this in three medical journals and in five top news venues that I respect. Among top news venues, one invited an op-ed, then turned it down after one week without any feedback. Conversely, The New York Times, offered multiple rounds of feedback over 8 days. Eventually, they rewrote the first half entirely, stated it will appear the next day, and then said they were sorry but they could not publish

the op-ed. STAT kept it for 5 days and sent extensive, helpful comments. I made extensive revisions, then they rejected it, apparently because an expert reviewer told them that "no infectious disease expert thinks this way" – paradoxically, I am trained and certified in infectious diseases.

Around 45%–53% of deaths in the US (Nursing Homes & Assisted Living Facilities Account for 45% of COVID-19 Deaths) (and as many or more in several European countries) (Danis et al., 2020) eventually were in nursing homes and related facilities, and probably another large share were nosocomial infections. An editor/reviewer at a top medical journal dismissed the possibility that many hospital staff are infected. Seroprevalence and PCR studies, however, have found very high infection rates in health care workers (Mansour et al., 2020; Sandri et al., 2020; Treibel T.A. Manisty, Burton, et al., 2020) and in nursing homes (Arons et al., 2020; Gandhi, Yokoe, & Havlir, 2020).

Had we dealt with this coronavirus considering what other widely-circulating coronaviruses do based on medical or infectious disease (not modeling) textbooks (=they cause mostly mild infections, but they can particularly devastate nursing homes and hospitals), (McIntosh, 2020; Patrick et al., 2006; Walsh et al., 2013) my foolish prediction might have been less ridiculous. Why was my article never accepted? Perhaps editors were influenced by some social media who painted me and my views as rather despicable and/or a product of "conservative ideology" (a stupendously weird classification, given my track record). As I say in my Stanford webpage: "I have no personal social media accounts - I admire people who can outpour their error-free wisdom in them, but I make a lot of errors, I need to revisit my writings multiple times before publishing, and I see no reason to make a fool of myself more frequently than it is sadly unavoidable". So, here I stand corrected.

References

- Actual Coronavirus Infections Vastly Undercounted, CDC. Data Shows - The New York Times, <https://www.nytimes.com/2020/06/27/health/coronavirus-antibodies-asymptomatic.html>. (Accessed 5 July 2020).
- AP counts: over 4500 virus patients sent to NY nursing homes (2020). Retrieved from <https://apnews.com/5ebc0ad45b73a899efa81f098330204c>. (Accessed 19 June 2020).
- Arons, M. M., Hatfield, K. M., Reddy, S. C., Kimball, A., James, A., et al. (2020). Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *New England Journal of Medicine*, 382, 2081–2090.
- Boccia, S., Ricciardi, W., & Ioannidis, J. P. A. (2020). What other countries can learn from Italy during the COVID-19 pandemic. *JAMA Internal Medicine*, <http://dx.doi.org/10.1001/jamainternmed.2020.1447>.
- Chin, V., Ioannidis, J. P. A., Tanner, M., & Cripps, S. (2020a). *Effects of non-pharmaceutical interventions on COVID-19: A Tale of Two Models*. medRxiv. <http://dx.doi.org/10.1101/2020.07.22.20160341>.
- Chin, V., Samia, N. I., Marchant, R., Rosen, O., Ioannidis, J. P. A., Tanner, M. A., et al. (2020b). A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York State. *European Journal of Epidemiology*, 35, 733–742.
- Cirillo, P., & Taleb, N. N. (2020). Tail risk of contagious diseases. *Nature Physics*, <http://dx.doi.org/10.1038/s41567-020-0921-x>.
- Danis, K., Fonteneau, L., Georges, S., et al. (2020). High impact of COVID-19 in long-term care facilities, suggestion for monitoring in the EU/EEA. *Euro Surveillance*, 25(22), pii=2000956. <http://dx.doi.org/10.2807/1560-7917.ES.2020.25.22.2000956>.

- De Filippo, O., D'Ascenzo, F., Angelini, F., et al. (2020). Reduced rate of hospital admissions for ACS during covid-19 outbreak in northern Italy. *The New England Journal of Medicine*, <http://dx.doi.org/10.1056/NEJMc2009166>.
- Ferguson, N. M., Donnelly, C. A., & Anderson, R. M. (2001a). The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science*, 292, 1155–1160.
- Ferguson, N. M., Donnelly, C. A., & Anderson, R. M. (2001b). Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature*, 413(6855), 542–548.
- Ferguson, N. M., Ghani, A. C., Donnelly, C. A., Hagensars, T. J., & Anderson, R. M. (2002). Estimating the human health risk from possible BSE infection of the British sheep flock. *Nature*, 415, 420–424.
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, <http://dx.doi.org/10.1038/s41586-020-2405-7>.
- Gandhi, M., Yokoe, D. S., & Havlir, D. V. (2020). Asymptomatic transmission the Achilles' Heel of current strategies to control covid-19. *New England Journal of Medicine*, 382, 2158–2160.
- Grisin, S. (2020;). Covid-19: Staggering number of extra deaths in community is not explained by covid-19. *The BMJ*, 369. <http://dx.doi.org/10.1136/bmj.m1931>.
- Hains, T. (2020). Ezekiel Emanuel: U.S. Will Have 100 Million Cases Of COVID-19 In Four Weeks, Doubling Every Four Days March 27, 2020. Retrieved from https://www.realclearpolitics.com/video/2020/03/27/ezekiel_emanuel_us_will_have_100_million_cases_of_covid-19_in_four_weeks.html. (Accessed 18 June 2020).
- Holmdahl, I., & Buckee, C. (2020). Wrong but useful – what Covid-19 epidemiologic models can and cannot tell us. *The New England Journal of Medicine*, (in press).
- Homburg, S., & Kuhbandner, C. (2020). Comment on Flaxman, et al. (2020), *Nature*, <http://dx.doi.org/10.1038/s41586-020-2405-7>: The illusory effects of non-pharmaceutical interventions on COVID-19 in Europe. In: <https://s3-eu-west-1.amazonaws.com/pstorage-advance-4589058904380/23182457/2020CommentFlaxmanPreprint.pdf>. (Accessed 20 June 2020).
- Huff, D., & Geis, I. (1954). *How to lie with statistics*. WW Norton & company.
- IHME COVID-19 health service utilization forecasting team, & Murray, Christopher J. L. (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. medRxiv 2020, <http://dx.doi.org/10.1101/2020.03.27.20043752>.
- Ioannidis, John P. A. Interview given in March and published April 1, 2020 in Der Spiegel: Ist das Coronavirus weniger tödlich als angenommen? Die drakonischen Maßnahmen, ...sind natürlich sinnvoll und auch gerechtfertigt, weil wir noch zu wenig über das Virus wissen (the draconian measures ... of course make perfect sense and they are appropriate, because we still know little about the virus. In: <https://www.spiegel.de/consent-a-?targetUrl=https%3A%2F%2Fwww.spiegel.de%2Fpolitik%2Fausland%2FCorona-krise-ist-das-virus-weniger-toedlich-als-angenommen-a-a6921df1-6e92-4f76-bddb-062d2bf7f441&ref=https%3A%2F%2Fwww.google.com%2F>. (Accessed 18 June 2020).
- Ioannidis, J. P. (2020). *The Totality of the Evidence*. Boston Review, Retrieved from <http://bostonreview.net/science-nature/john-p-ioannidis-totality-evidence>. (Accessed 2 June 2020).
- Ioannidis, J. P. (2020a). *A fiasco-in-the-making? As the Coronavirus pandemic takes hold we are making decisions without reliable data*. STAT, Retrieved from <https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>. (Accessed 18 June 2020).
- Ioannidis, John P. A. (2020). Interview given for Journeyman Pictures on March 24, uploaded as detailed 1-hour video, Perspectives on the Pandemic, episode 1. I am perfectly happy to be in a situation of practically lockdown in California more or less, with shelter-in-place, but I think very soon we need to have that information to see what we did with that and where do we go next. The video had over 1 million views but was censored in May by YouTube.
- Ioannidis, John P. A. (2020c). Interview given on March 22 to in2life.gr: Σωστή η καραντίνα, αλλά... (“The lockdown is the correct decision, but...”): Τα αυστηρά μέτρα που παίρνονται σε αυτή τη φάση είναι απαραίτητα επειδή φαίνεται ότι ο έλεγχος της επιδημίας έχει ξεφύγει (the stringent measures that are taken at this phase are necessary because it seems that we have lost control of the epidemic, <https://www.in2life.gr/features/notes/article/1004752/ioannhs-ioannidhs-sosth-tora-h-karantina-alla-.htm>. (Accessed 18 June 2020).
- Ioannidis, J. P. (2020d). Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. *European Journal of Clinical Investigation*, 50, Article e13222.
- Ioannidis, J. P. A., Axford, C. A., & Contopoulos-Ioannidis, J. P. (2020). Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environmental Research*, 188, article 890.
- Jefferson, T., Del Mar, C. B., Dooley, L., et al. (2011). Physical interventions to interrupt or reduce the spread of respiratory viruses. *The Cochrane Database of Systematic Reviews*, (7), CD006207.
- Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA*, <http://dx.doi.org/10.1001/jama.2020.6585>, Epub 2020/04/17.
- Kitching, R. P., Thrusfield, M. V., & Taylor, N. M. (2006). Use and abuse of mathematical models: an illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Revue Scientifique et Technique*, 25, 293–311.
- Mansour, M., Leven, E., Muellers, K., Stone, K., Mendu, D. R., & Wainberg, A. (2020). A Prevalence of SARS-CoV-2 antibodies among healthcare workers at a tertiary academic hospital in New York City. *Journal of General Internal Medicine*, 3, 1–2. <http://dx.doi.org/10.1007/s11606-020-05926-8>.
- Marchant, R., Samia, N. I., Rosen, O., Tanner, M. A., & Cripps, S. (2020). Learning as we go: An examination of the statistical accuracy of COVID19 daily death count predictions. medRxiv 2020 <http://dx.doi.org/10.1101/2020.04.11.20062257>.
- McIntosh, K. (2020). Coronaviruses. UpToDate. <https://www.uptodate.com/contents/coronaviruses>. (Accessed 19 June 2020).
- Melnick, T., & Ioannidis, J. P. (2020). Ioannidis JP should governments continue lockdown to slow the spread of covid-19?. *BMJ*, 369, m1924.
- Moser, D. A., Glaus, J., Frangou, S., et al. (2020). Years of life lost due to the psychosocial consequences of COVID19 mitigation strategies based on Swiss data. *European Psychiatry*, 1–14.
- National Highway Traffic Safety Administration (2017). *Lives saved in 2016 by restraint use and minimum-drinking-age laws*. Washington, DC: US Department of Transportation, Publication no. DOT-HS-812-454. Available at <https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812454External>. (Accessed 20 June 2020).
- Nursing Homes & Assisted Living Facilities Account for 45% of COVID-19 Deaths, (2020). Retrieved from <https://freopp.org/the-covid-19-nursing-home-crisis-by-the-numbers-3a47433c3f70>. (Accessed 5 July 2020).
- Patrick, D. M., Petric, M., Skowronski, D. M., et al. (2006). An outbreak of human coronavirus OC43 infection and serological cross-reactivity with SARS coronavirus. *Canadian Journal of Infectious Diseases and Medical Microbiology*, 17(6), 330.
- Roxby, A. C., Greninger, A. L., Hatfield, K. M., Lynch, J. B., Dellit, T. H., James, A., et al. (2020). Detection of SARS-CoV-2 Among residents and staff members of an independent and assisted living community for older adults - Seattle, Washington, 2020. *Morbidity and Mortality Weekly Report*, 69(14), 416–418.
- Saltelli, A., Bamber, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., et al. (2020). Five ways to ensure that models serve society: a manifesto. *Nature*, 582, 482–484.
- Sandri, M. T., Azzolini, E., Torri, V., Carloni, S., Tedeschi, M., Castoldi, M., et al. (2020). IgG serology in health care and administrative staff populations from 7 hospital representative of different exposures to SARS-CoV-2 in lombardy, Italy. medRxiv. <http://dx.doi.org/10.1101/2020.05.24.20111245>.

- Saunders-Hastings, P., Crispo, J. A. G., Sikora, L., & Krewski, D. (2017). Effectiveness of personal protective measures in reducing pandemic influenza transmission: a systematic review and meta-analysis. *Epidemics*, 20, 1–20.
- Sud, A., Jones, M., Broggio, J., Loveday, C., Torr, B., Garrett, A., et al. (2020). Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic. *Annals of Oncology*, 16, S0923-7534(20)39825-2. <http://dx.doi.org/10.1016/j.annonc.2020.05.009>.
- Taleb, N. N. (2020). *On single point forecasts for fat tailed variables*. International Institute of Forecasters, Retrieved from: <https://forecasters.org/blog/2020/06/14/on-single-point-forecasts-for-fat-tailed-variables/>.
- Tanner, M. A. (1996). *Tools for statistical inference*. Springer.
- Toker, D., Sommer, F. T., & D'Esposito, M. (2000). A simple method for detecting chaos in nature. *Communications Biology*, 3, 11. <http://dx.doi.org/10.1038/s42003-019-0715-9>.
- Treibel T.A. Manisty, C., Burton, M., et al. (2020). Covid-19: PCR screening of asymptomatic health-care workers at London hospital. *Lancet*, 395, 1608–1610.
- UK government (2009). The 2009 influenza pandemic review. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/61252/the2009influenzapandemic-review.pdf. (Accessed 2 June 2020).
- US could see 100,000 new COVID-19 cases per day, Fauci says, (2020). Retrieved from <https://www.statnews.com/2020/06/30/u-s-could-see-100000-new-covid-19-cases-per-day-fauci-says/>. (Accessed 5 July 2020).
- Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., et al. (2005). Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *Journal of Virology*, 79, 1595–1604.
- Walsh, E. E., Shin, J. H., & Falsey, A. R. (2013). Clinical impact of human coronaviruses 229e and oc43 infection in diverse adult populations. *Journal of Infectious Diseases*, 208, 1634–1642.
- Williamson, E., Walker, A. J., Bhaskaran, K. J., Bacon, S., Bates, C., Morton, C. E., et al. (2020). OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. 584, 430–6,