

Forecasting for COVID-19 has failed

John P.A. Ioannidis (1), Sally Cripps (2), Martin A. Tanner (3)

(1) 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, California, USA

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

(3) Department of Statistics, Northwestern University, Evanston, Illinois, USA

Correspondence to: John P.A. Ioannidis, MD, DSc, Stanford Prevention Research Center, 1265

Welch Road, Medical School Office Building, Room X306, USA. E-mail:

jioannid@stanford.edu

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, looking at 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.

1. Initial position

COVID-19 is a major acute crisis with unpredictable consequences. Many scientists have struggled to make forecasts about its impact.¹ However, despite involving many excellent modelers, best intentions, and highly sophisticated tools, forecasting efforts have largely failed.

Experienced modelers drew early on parallels between COVID-19 and the Spanish flu² that caused >50 million deaths with mean age at death being 28. We all lament the current loss of life. However, as of June 18, 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.³ 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).⁴ However, eventually very few hospitals were stressed, for a couple of weeks. Most hospitals maintained largely empty wards, expecting tsunamis that never came. The general population 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,⁵ important treatments (e.g. for cancer) were unjustifiably delayed,⁶ mental health suffered.⁷ With damaged operations, many hospitals started losing personnel, reducing capacity to face future crises (e.g. a second wave). With massive new unemployment, more people may lose health insurance. Prospects of starvation and of lack of control for other infectious diseases (like tuberculosis, malaria, and childhood communicable diseases where vaccination is hindered by COVID-19 measures) are dire.^{8,9}

Modeling resurgence after reopening also failed (Table 2). E.g. a Massachusetts General Hospital model¹⁰ predicted over 23,000 deaths within a month of Georgia reopening – 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, e.g. using early SIR-based models to predict what would happen in the entire season. However, even forecasting built directly on data alone fared badly,^{11,12} failing not only in ICU bed predictions (Figure 1) but even in *next day* death predictions when issues of long-term chaotic behavior do not come into play (Figures 2 and 3). Even for short-term forecasting when the epidemic wave waned, models presented confusingly diverse predictions with huge uncertainty (Figure 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.¹³ Eventually 457 deaths occurred.¹⁴ Models on foot-and-mouth disease by top scientists in top journals^{15,16} were subsequently questioned¹⁷ 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.¹⁸ However, the lower bound predicted as low as 50 deaths,¹⁸ 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. Actually, erroneous predictions may have even been useful. A wrong, doomsday prediction may incentivize people

1
2
3
4 towards better personal hygiene. Problems start when public leaders take (wrong) predictions too
5
6 seriously, considering them crystal balls without understanding their uncertainty and the
7
8 assumptions made. Slaughtering millions of animals may aggravate animal business stakeholders
9
10 – but most citizens are not directly affected. However, with COVID-19, espoused wrong
11
12 predictions can devastate billions of people in terms of the economy, health, and societal turmoil
13
14 at-large.
15
16

17
18 Let's be clear: even if millions of deaths did not happen this season, they may happen
19
20 with the next wave, next season, or some new virus in the future. A doomsday forecast may
21
22 come handy to protect civilization, when and if calamity hits. However, even then, we have little
23
24 evidence that aggressive measures focusing only on few dimensions of impact actually reduce
25
26 death toll and do more good than harm. We need models which incorporate multicriteria
27
28 objective functions. Isolating infectious impact, from all other health, economic and social
29
30 impacts is dangerously narrow-minded. More importantly, with epidemics becoming easier to
31
32 detect, opportunities for declaring global emergencies will escalate. Erroneous models can
33
34 become powerful, recurrent disruptors of life on this planet. Civilization is threatened from
35
36 epidemic incidentalomas.
37
38
39
40
41

42
43 Cirillo and Taleb thoughtfully argue¹⁹ that when it comes to contagious risk, we should
44
45 take doomsday predictions seriously: major epidemics follow a fat-tail pattern and extreme value
46
47 theory becomes relevant. Examining 72 major epidemics recorded through history, they
48
49 demonstrate a fat-tailed mortality impact. However, they analyze only the 72 most-noticed
50
51 outbreaks, a sample with astounding selection bias. For example, according to their dataset, the
52
53 first epidemic originating from sub-Saharan Africa did not occur until 1920 AD, namely
54
55 HIV/AIDS. The most famous outbreaks in human history are preferentially selected from the
56
57
58
59
60
61
62
63
64
65

1
2
3
4 extreme tail of the distribution of all outbreaks. Tens of millions of outbreaks with a couple
5
6 deaths must have happened throughout time. Probably hundreds of thousands might have
7
8 claimed dozens of fatalities. Thousands of outbreaks might have exceeded 1,000 fatalities. Most
9
10 eluded the historical record. The four garden variety coronaviruses may be causing such
11
12 outbreaks every year.^{20,21} One of them, OC43 seems to have been introduced in humans as
13
14 recently as 1890, probably causing a “bad influenza year” with over a million deaths.²² Based on
15
16 what we know now, SARS-CoV-2 may be closer to OC43 than SARS-CoV-1. This does not
17
18 mean it is not serious: its initial human introduction can be highly lethal, unless we protect those
19
20 at risk.
21
22
23
24

25
26 A heavy tail distribution ceases to be as heavy as Taleb imagines when the middle of the
27
28 distribution becomes much larger. One may also argue that pandemics, as opposed to epidemics
29
30 without worldwide distribution, are more likely to be heavy-tailed. However, the vast majority of
31
32 the 72 contagious events listed by Taleb were not pandemics, but localized epidemics with
33
34 circumscribed geographic activity. Overall, when a new epidemic is detected, it is even difficult
35
36 to pinpoint which distribution of which known events it should be mapped against.
37
38
39
40

41 Blindly acting based on extreme value theory alone would be sensible if we lived in the
42
43 times of the Antonine plague or even in 1890, with no science to identify the pathogen, elucidate
44
45 its true prevalence, estimate accurately its lethality, and carry out good epidemiology to identify
46
47 which people and settings are at risk. Until we accrue this information, immediate better-safe-
48
49 than-sorry responses are legitimate, trusting extreme forecasts as possible (not necessarily likely)
50
51 scenarios. However, caveats of these forecasts should not be ignored^{1,23} and new evidence on the
52
53 ground truth needs continuous reassessment. Upon acquiring solid evidence about the
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 epidemiological features of new outbreaks, implausible, exaggerated forecasts²⁴ should be
5
6 abandoned. Otherwise, they may cause more harm than the virus itself.
7
8

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

10
11 The insightful recent essay of Taleb²⁵ offers additional opportunities for fruitful
12
13 discussion.
14
15

16 *2.1 Point estimate predictions and technical points*

17
18 Taleb²⁵ ruminates on the point of making point predictions. Serious modelers (whether
19
20 frequentist or Bayesian) would never rely on point estimates to summarize skewed distributions.
21
22 Even an early popular presentation²⁶ from 1954 has a figure (see page 33) with striking
23
24 resemblance to Taleb's Figure 1.²⁵ In a Bayesian framework, we rely on the full posterior
25
26 predictive distribution, not single points.²⁷ Moreover, Taleb's choice of a three-parameter Pareto
27
28 distribution is peculiar. It is unclear this model provides a measurably better fit to his (hopelessly
29
30 biased) pandemic data¹⁹ than, say, a two parameter Gamma distribution fitted to log counts.
31
32 Regardless, either skewed distribution would then have to be modified to allow for the use of all
33
34 available sources of information in a logically consistent fully probabilistic model, e.g. via a
35
36 Bayesian hierarchical model (which can certainly be formulated to accommodate fat tails if
37
38 needed). In this regard, we note that examining the NY daily death count data studied in ref. 12,
39
40 these data are found to be characterized as stochastic rather than chaotic.²⁸ Taleb seems to fit an
41
42 unorthodox model, and then abandons all effort to predict anything. He simply assumes
43
44 doomsday has come, much like a panic-driven Roman would have done in the Antonine plague,
45
46 lacking statistical, biological, and epidemiological insights.
47
48
49
50
51
52
53
54

55 *2.2. Should we wait for the best evidence before acting?*

56
57
58
59
60
61
62
63
64
65

1
2
3
4 Taleb²⁵ caricatures the position of a hotly debated mid-March op-ed by one of us,²⁹
5
6 alluring it “made statements to the effect that one should wait for “more evidence” before acting
7
8 with respect to the pandemic”, an obvious distortion of the op-ed. Anyone who reads the op-ed
9
10 unbiasedly realizes that it says exactly the opposite. It starts with the clear, unquestionable
11
12 premise that the pandemic is taking hold and is a serious threat. Immediate lockdown certainly
13
14 makes sense when an estimated 50 million deaths are possible. This was stated emphatically in
15
16 multiple occasions these days in interviews in multiple languages -for examples see refs. 30-32.
17
18 Certainly, adverse consequences of short-term lockdown cannot match 50 million lives.
19
20 However, better data can recalibrate estimates, re-assessing downstream the relative balance of
21
22 benefits and harms of longer-term prolongation of lockdown. That re-appraised balance changed
23
24 markedly over time.⁹
25
26
27
28
29
30

31 Another gross distortion propagated in social media is that supposedly the op-ed²⁹ had
32
33 predicted that only 10,000 deaths in the USA. The key message of the op-ed was that we lack
34
35 reliable data, i.e. we don’t know. The self-contradicting misinterpretation as “we don’t know, but
36
37 actually we do know that 10,000 deaths will happen” is impossible. The op-ed discussed two
38
39 extreme scenarios to highlight the tremendous uncertainty absent reliable data: 10,000 deaths in
40
41 the US and 40,000,000 deaths. We needed reliable data, quickly, to narrow this vast uncertainty.
42
43 We did get data and did narrow uncertainty. Science did work eventually, even if forecasts,
44
45 including those made by one of us (confessed and discussed in Box 1), failed.
46
47
48
49

50 51 *2.3 Improper and proper analogies of benefit-risk*

52
53 Taleb²⁵ offers several analogies to assert that all precautionary actions are justified in
54
55 pandemics, deriding “waiting for the accident before putting the seat belt on, or evidence of fire
56
57 before buying insurance”.²⁵ The analogies assume that the cost of precautionary actions are small
58
59
60
61
62
63
64
65

1
2
3
4 in comparison to the cost of the pandemic, and that the consequences of the action have little
5
6 impact on it. However, precautionary actions can backfire badly when they are misinformed. In
7
8 March, modelers were forecasting collapsed health systems, e.g. 140,000 beds would be needed
9
10 in New York, when only a small fraction were available. Precautionary actions damaged the
11
12 health system, increased COVID-19 deaths,³³ and exacerbated other health problems (Table 4).
13
14

15
16 Seat belts cost next to nothing to produce in cars and have unquestionable benefits.
17
18 Despite some risk compensation and some excess injury with improper use, eventually seat belts
19
20 prevent ~50% of serious injuries and deaths.³⁴ Measures for pandemic prevention equivalent to
21
22 seat belts in terms of benefit-harm profile are simple interventions like hand washing, respiratory
23
24 etiquette and mask use in appropriate settings: large proven benefit, no/little harm/cost.^{35,36} Even
25
26 before the COVID-19 pandemic, we had randomized trials showing 38% reduced odds of
27
28 influenza infection with hand washing and (non-statistically significant, but possible) 47%
29
30 reduced odds with proper mask wearing.³⁵ Despite lack of trials, it is sensible and minimally
31
32 disruptive to avoid mass gatherings and decrease unnecessary travel. Prolonged draconian
33
34 lockdown is not equivalent to seat belts. It resembles forbidding all commute.
35
36
37
38
39
40

41 Similarly fire insurance offers a misleading analogy. Fire insurance makes sense only at
42
43 reasonable price. Draconian prolonged lockdown may be equivalent to paying fire insurance at a
44
45 price higher than the value of the house.
46
47

48 *2.4 Mean, observed maximum, and more than the observed maximum*

49

50
51 Taleb refers to the Netherlands where maximum values for flooding, not the mean, are
52
53 considered.²⁵ Anti-flooding engineering has substantial cost, but a favorable decision-analysis
54
55 profile after considering multiple types of impact. Lockdown measures were decided based on
56
57 examining only one type of impact, COVID-19. Moreover, the observed flooding maximum to-
58
59
60
61
62
63
64
65

1
2
3
4 date does not preclude even higher future values. Netherlands aims to avoid devastation from
5
6 floods occurring once every 10,000 years in densely populated areas.³⁷ A more serious flooding
7
8 event (e.g. one that occurs every 20,000 years) may still submerge the Netherlands next week.
9
10 However, prolonged total lockdown is not equivalent to building higher sea walls. It is more like
11
12 abandoning the country - asking the Dutch to immigrate, because their land is quite unsafe.
13

14
15 Other natural phenomena also exist where high maximum risks are difficult to pinpoint
16
17 and where new maxima may be reached. E.g., following Taleb's argumentation, one should
18
19 forbid living near active volcanoes. Living at the Santorini caldera is not exciting, but foolish:
20
21 that dreadful island should be summarily evacuated. Same applies to California: earthquake
22
23 devastation may strike any moment. Prolonged lockdown zealots might barely accept a
24
25 compromise: whenever substantial seismic activity occurs, California should be temporarily
26
27 evacuated until all seismic activity ceases.
28
29
30
31
32

33
34 Furthermore, fat-tailed uncertainty and approaches based on extreme value theory may be
35
36 useful before a potentially high-risk phenomenon starts and during its early stages. However, as
37
38 more data accumulate and the high-risk phenomenon can be understood more precisely with
39
40 plenty of data, the laws of large numbers may apply and stochastic rather than chaotic
41
42 approaches may become more relevant and useful than continuing to assume unlikely extremes.
43
44 Further responses to Taleb²⁵ appear in Table 5.
45
46

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

48 *3.1 How do we move forward to deal with the COVID-19 threat?*

49
50 The short answer is: using science and more reliable data. We can choose measures with
51
52 favorable benefit-risk ratio, when we consider together multiple types of impact, not only on
53
54 COVID-19, but on health as a whole, as well as society and economy.
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 Currently we know that approximately half of the COVID-19 deaths in Europe and the
5
6 USA affected nursing home residents.^{38,39} Another sizeable proportion were nosocomial
7
8 infections.⁴⁰ If we protect these locations with draconian hygiene measures and intensive testing,
9
10 we may avert 70% of the fatalities without large-scale societal disruption and without adverse
11
12 consequences on health. Other high-risk settings, e.g. prisons, homeless shelters, meat-
13
14 processing plants also need aggressive protection. For the rest of the population, we have strong
15
16 evidence on a very steep age gradient with ~1000-fold differences in death risk for people >80
17
18 years old versus children.⁴¹ We have also detailed insights on how different background diseases
19
20 modify COVID-19 risk for death or other serious outcome.⁴² We can use hygiene and some least
21
22 disruptive distancing measures to protect people. We can use intensive testing (i.e. again, use
23
24 science) to detect resurgence of epidemic activity and extinguish it early – the countries that
25
26 faced most successfully the first wave, e.g. Singapore and Taiwan, did exactly that highly
27
28 successfully. We can use data to track how the epidemic and its impact evolves. Data can help
29
30 inform more granular models and titrate decisions considering distributions of risk (Figure 5).⁴²

3.2 *Abandon or improve epidemic forecasting?*

41 Poorly performing models and models that perform well for only one dimension of
42
43 impact can cause harm. It is not just an issue of academic debate, it is an issue of potentially
44
45 devastating, wrong decisions.³⁶ Taleb²⁵ seems self-contradicting: does he espouse abandoning all
46
47 models (since they are so wrong) or using models but always assuming the worst? However,
48
49 there is no single worst scenario, but a centile of the distribution: should we prepare for an event
50
51 that has 0.1%, 0.001%, or 0.000000000001% chance of happening? Paying what price in harms?
52
53

54 Abandoning all epidemic modeling appears too unrealistic. Besides identifying the
55
56 problems of epidemic modeling, Table 3 also offers suggestions on addressing some of them.
57
58
59
60
61
62
63
64
65

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

- Invest more on collecting, cleaning, and curating real, unbiased data, 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.
- Incorporate best epidemiological estimates on age structure and comorbidities in the modeling
- Focus on quality-adjusted life-years rather than deaths
- Avoid unrealistic assumptions about benefits of interventions; don't 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 make sure 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

1
2
3
4 Of interest, another group of researchers have reached⁴³ almost identical conclusions and
5
6 their recommendations are largely overlapping. Forecasting is dangerous to hype.
7
8

9 Importantly, not all problems can be fixed. At best, epidemic models offer only tentative
10
11 evidence. Great caution and nuance are still needed. Models that use reliable data, that are
12
13 validated and continuously reappraised for their performance in real-time, and that combine
14
15 multiple dimensions of impact may have more utility. A good starting point is to acknowledge
16
17 that problems exist. Serious scientists who have published poorly performing models should
18
19 acknowledge this. They may also correct or even retract their papers, getting credit and
20
21 congratulations, not blame, for corrections/retractions. The worst nightmare would be if
22
23 scientists and journals insist that prolonged draconian measures cause the massive difference
24
25 between predictions and eventual outcomes. Serious scientists and serious journals⁴⁴ are
26
27 unfortunately flirting with this slippery, defensive path.⁴⁵ Total lockdown is a bundle of dozens
28
29 of measures. Some may be very beneficial, but some others may be harmful. Hiding uncertainty
30
31 can cause major harm downstream and leaves us unprepared for the future. For papers that fuel
32
33 policy decisions with major consequences, transparent availability of data, code, and named
34
35 peer-review comments is also a minimum requirement.
36
37
38
39
40
41
42

43 The possibility of calibrating model predictions for looking at extremes rather than just
44
45 means is sensible, especially in early days of pandemics, when much is unknown about the virus
46
47 and its epidemiological footprint. However, when calibration/communication on extremes is
48
49 adopted, one should also consider similar calibration for the potential harms of adopted
50
51 measures. For example, tuberculosis has killed 1 billion people in the last 200 years, it still kills
52
53 1.5 million people (mostly young and middle age ones) annually, and prolonged lockdown may
54
55 cause 1.4 million extra tuberculosis deaths between 2020-2025.⁴⁶ Measles has killed about 200
56
57
58
59
60
61
62
63
64
65

1
2
3
4 million people in the last 150 years; disrupted MMR vaccination may fuel lethal recrudescence.⁴⁷
5
6 Use of extreme case predictions for COVID-19 deaths should be co-examined with extreme case
7
8 predictions for deaths and impact from many other lockdown-induced harms. Models should
9
10 provide the big picture of multiple dimensions. Similar to COVID-19, as more reliable data
11
12 accrue, predictions on these other dimensions should also be corrected accordingly.
13
14

15
16 Eventually, it is probably impossible (and even undesirable) to ostracize epidemic
17
18 forecasting, despite its failures. Arguing that forecasting for COVID-19 has failed should not be
19
20 misconstrued to mean that science has failed. Developing models in real time for a novel virus,
21
22 with poor quality data, is a formidable task and the groups who attempted this and made public
23
24 their predictions and data in a transparent manner should be commended. We readily admit that
25
26 it is far easier to criticize a model than to build one. It would be horrifically retrograde if this
27
28 debate ushers in a return to an era where predictions, on which huge decisions are made, are kept
29
30 under lock and key (e.g. by the government - as is the case in Australia).
31
32

33 34 35 *3.3. Learning from the COVID-19 pandemic and from our mistakes* 36 37

38 We wish to end on a more positive note, namely where we feel forecasting has been
39
40 helpful. Perhaps the biggest contribution of these models is that they serve as a springboard for
41
42 discussions and debates. Dissecting variation in performance of various models (e.g. casting a
43
44 sharp eye to circumstances where a particular model excelled) can be highly informative and a
45
46 systematic approach to the development and evaluation of such models is needed.¹² This
47
48 demands a coherent approach to collecting, cleaning and curating data, as well as a transparent
49
50 approach to evaluating the suitability of models with regard to predictions and forecast
51
52 uncertainty.
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 What we have learned from the COVID-19 pandemic can be passed to future generations
5
6 that hopefully should be better prepared to deal with a new, different pandemic, learning from
7
8 our failures. There is no doubt that, again, an explosive literature of models and forecasting will
9
10 emerge again as soon as a new pandemic is suspected. However, we can learn from our current
11
12 mistakes to be more cautious with interpreting, using, and optimizing these models. Being more
13
14 cautious does not mean not to act decisively, but it requires looking at the totality of the data;
15
16 considering multiple types of impact; having scientists from very different disciplines involved;
17
18 replacing speculations, theories and assumptions with real, empirical data as quickly as possible;
19
20 and modifying and aligning decisions to the evolving best evidence.
21
22
23
24

25
26 In the current pandemic, we largely failed to protect people and settings at risk. We could
27
28 have done much better in this regard. It is difficult to correct mistakes that have already led to
29
30 people dying, but we can avoid making the same mistakes in future pandemics from different
31
32 pathogens. We can avoid making the same mistakes even for COVID-19 going forward, since
33
34 this specific pandemic has not ended as we write. In fact, its exact eventual impact is still
35
36 unknown. For example, the leader of the US task force, Dr. Anthony Fauci, recently warned of
37
38 reaching 100,000 COVID-19 US cases per day.⁴⁸ Maybe this prediction is already an
39
40 underestimate, because with over 50,000 cases diagnosed per day in early July 2020, the true
41
42 number of infections may be many times larger. There is currently wide agreement that the
43
44 number of infections in many parts of the United States is more than 10 times higher than the
45
46 reported rates.⁴⁹ We do have interventions that can prevent or reduce the resurgence of the
47
48 epidemic wave. Moreover, we know that 100,000 cases in healthy children and young adults
49
50 may translate to almost 0 deaths. Conversely, 100,000 cases in high-risk susceptible individuals
51
52 and settings may translate to many thousands of deaths. We can use science to extinguish
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 epidemic waves, or, if this is impossible, have them spend their flare on settings where they carry
5
6 minimal risk. The same forecast for the number of cases may vary 1000-fold or more in terms of
7
8 outcomes that matter. We should use forecasting, along with many other tools and various types
9
10 of evidence to improve outcomes that matter.
11
12
13
14
15
16
17
18

19 Acknowledgement: The authors thank Vincent Chin helpful discussions and for providing
20
21 figures 1-3.
22
23
24
25

26 Conflicts of interest: none
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

1. Holmdahl I, Buckee C. Wrong but useful — what Covid-19 epidemiologic models can and cannot tell us. N Engl J Med 2020 (in press).
2. <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/>, last accessed June 2, 2020.
3. Hains, T. “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, last accessed: June 18, 2020.
4. IHME COVID-19 health service utilization forecasting team, Christopher JL Murray. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. medRxiv 2020, doi: <https://doi.org/10.1101/2020.03.27.20043752>.
5. De Filippo O, D’Ascenzo F, Angelini F, et al. Reduced rate of hospital admissions for ACS during covid-19 outbreak in northern Italy. N Engl J Med 2020. doi:10.1056/NEJMc2009166.
6. Sud A, Jones M, Broggio J, Loveday C, Torr B, Garrett A, Nicol DL, Jhanji S, Boyce SA, Gronthoud F, Ward P, Handy JM, Yousaf N, Larkin J, Suh YE, Scott S, Pharoah PDP, Swanton C, Abbosh C, Williams M, Lyratzopoulos G, Houlston R, Turnbull C. Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic. Ann Oncol. 2020 May 16:S0923-7534(20)39825-2. doi: 10.1016/j.annonc.2020.05.009.

7. Moser DA, Glaus J, Frangou S, et al. Years of life lost due to the psychosocial consequences of COVID19 mitigation strategies based on Swiss data. *European Psychiatry* 2020:1-14.
8. Melnick T, Ioannidis JP. Should governments continue lockdown to slow the spread of covid-19? *BMJ* 2020;369:m1924.
9. Ioannidis JP. The totality of the evidence. *Boston Review*. In: <http://bostonreview.net/science-nature/john-p-ioannidis-totality-evidence>, last accessed June 2, 2020.
10. In: <https://www.massgeneral.org/news/coronavirus/COVID-19-simulator>, last accessed June 2, 2020.
11. Marchant R, Samia NI, Rosen O, Tanner MA, Cripps S. Learning as We Go: An Examination of the Statistical Accuracy of COVID19 Daily Death Count Predictions. *medRxiv* 2020 doi: <https://doi.org/10.1101/2020.04.11.20062257>
12. Chin V, Samia NI, Marchant R, Rosen O, Ioannidis JPA, Tanner MA and Cripps S. A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York State (submitted).
13. <https://www.theguardian.com/uk/2009/jul/16/swine-flu-cases-rise-britain>, last accessed June 2, 2020.
14. UK government. The 2009 influenza pandemic review. In: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/61252/the2009influenzapandemic-review.pdf. Last accessed June 2, 2020

- 1
2
3
4 15. Ferguson NM, Donnelly CA, Anderson RM. Transmission intensity and impact of
5
6 control policies on the foot and mouth epidemic in Great Britain. *Nature*
7
8 2001;413(6855):542-8.
9
10
11 16. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great
12
13 Britain: Pattern of spread and impact of interventions. *Science* 2001;292:1155–60.
14
15
16 17. Kitching RP, Thrusfield MV, Taylor NM. Use and abuse of mathematical models: an
17
18 illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Rev*
19
20 *Sci Tech* 2006;25:293-311.
21
22
23 18. Ferguson NM, Ghani AC, Donnelly CA, Hagens TJ, Anderson RM. Estimating the
24
25 human health risk from possible BSE infection of the british sheep flock. *Nature*
26
27 2002;415:420-4.
28
29
30 19. Cirillo P, Taleb NN. Tail risk of contagious diseases. *Nature Physics* 2020;
31
32 <https://doi.org/10.1038/s41567-020-0921-x>
33
34
35 20. Walsh EE, Shin JH, Falsey AR. Clinical impact of human coronaviruses 229E and OC43
36
37 infection in diverse adult populations. *J Infect Dis* 2013;208:1634-42.
38
39
40 21. Patrick DM, Petric M, Skowronski DM, et al. An outbreak of human coronavirus OC43
41
42 infection and serological cross-reactivity with SARS coronavirus. *Can J Infect Dis Med*
43
44 *Microbiol* 2006;17(6):330.
45
46
47 22. Vijgen L, Keyaerts E, Moës E, Thoelen I, Wollants E, Lemey P, et al. Complete genomic
48
49 sequence of human coronavirus OC43: molecular clock analysis suggests a relatively
50
51 recent zoonotic coronavirus transmission event. *J Virol.* 2005;79:1595–1604.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

23. Jewell NP, Lewnard JA, Jewell BL. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. JAMA. 2020. Epub 2020/04/17. doi: 10.1001/jama.2020.6585.
24. Ioannidis JP. Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. Eur J Clin Invest 2020;50:e13222.
25. Taleb NN. On single point forecasts for fat tailed variables. International Institute of Forecasters, 11 June 2020, available at: <https://forecasters.org/blog/2020/06/14/on-single-point-forecasts-for-fat-tailed-variables/>
26. Huff D, Geis I. "How to Lie with Statistics", WW Norton & company, 1954.
27. Tanner MA. Tools for Statistical Inference. Springer, 1996.
28. Toker D, Sommer FT, D'Esposito M. A simple method for detecting chaos in nature. Commun Biol 2000;3:11. doi.org/10.1038/s42003-019-0715-9
29. Ioannidis JP. A fiasco-in-the-making? As the coronavirus pandemic takes hold we are making decisions without reliable data. STAT, March 17, 2020, in: <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/>, last accessed June 18, 2020.
30. John P.A. Ioannidis. 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 You Tube.

- 1
2
3
4 31. John P.A. Ioannidis. Interview given on March 22 to in2life.gr: “Σωστή η καραντίνα,
5
6 αλλά...” (“The lockdown is the correct decision, but...”): “Τα αυστηρά μέτρα που
7
8 παίρνονται σε αυτή τη φάση είναι απαραίτητα επειδή φαίνεται ότι ο έλεγχος της
9
10 επιδημίας έχει ξεφύγει” (“the stringent measures that are taken at this phase are necessary
11
12 because it seems that we have lost control of the epidemic”,
13
14 [https://www.in2life.gr/features/notes/article/1004752/ioannhs-ioannidhs-sosth-tora-h-](https://www.in2life.gr/features/notes/article/1004752/ioannhs-ioannidhs-sosth-tora-h-karantina-alla-.htm)
15
16 [karantina-alla-.htm](https://www.in2life.gr/features/notes/article/1004752/ioannhs-ioannidhs-sosth-tora-h-karantina-alla-.htm), last accessed June 18, 2020.
17
18
19
20
21 32. John P.A. Ioannidis. Interview given in March and published April 1, 2020 in Der
22
23 Spiegel: Ist das Coronavirus weniger tödlich als angenommen? “Die drakonischen
24
25 Maßnahmen, ...sind natürlich sinnvoll und auch gerechtfertigt, weil wir noch zu wenig
26
27 über das Virus wissen” (“the draconian measures ... of course make perfect sense and
28
29 they are appropriate, because we still know little about the virus”. In:
30
31 [https://www.spiegel.de/consent-a-](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)
32
33 [?targetUrl=https%3A%2F%2Fwww.spiegel.de%2Fpolitik%2Fausland%2Fcorona-krise-](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)
34
35 [ist-das-virus-weniger-toedlich-als-angenommen-a-a6921df1-6e92-4f76-bddb-](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)
36
37 [062d2bf7f441&ref=https%3A%2F%2Fwww.google.com%2F](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), last accessed June 18,
38
39
40
41
42
43 2020.
44
45
46 33. AP counts: over 4500 virus patients sent to NY nursing homes.
47
48 <https://apnews.com/5ebc0ad45b73a899efa81f098330204c>, last accessed June 19, 2020.
49
50
51 34. National Highway Traffic Safety Administration. Lives saved in 2016 by restraint use
52
53 and minimum-drinking-age laws. US Department of Transportation, Washington, DC;
54
55 2017. Publication no. DOT-HS-812-454. Available at
56
57
58
59
60
61
62
63
64
65

- https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812454External. Last accessed June 20, 2020.
35. Saunders- Hastings P, Crispo JAG, Sikora L, Krewski D. Effectiveness of personal protective measures in reducing pandemic influenza transmission: a systematic review and meta- analysis. *Epidemics*. 2017;20:1- 20.
36. Jefferson T, Del Mar CB, Dooley L, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database Syst Rev*. 2011;(7):CD006207.
37. https://en.wikipedia.org/wiki/Flood_control_in_the_Netherlands, last accessed June 18, 2020.
38. Nursing Homes & Assisted Living Facilities Account for 45% of COVID-19 Deaths. <https://freopp.org/the-covid-19-nursing-home-crisis-by-the-numbers-3a47433c3f70>, last accessed July 5, 2020.
39. Danis K, Fonteneau L, Georges S, et al. High impact of COVID-19 in long-term care facilities, suggestion for monitoring in the EU/EEA, May 2020. *Euro Surveillance* 2020;25(22):pii=2000956. <https://doi.org/10.2807/1560-7917.ES.2020.25.22.200095>
40. Boccia S, Ricciardi W, Ioannidis JPA. What other countries can learn from Italy during the COVID-19 pandemic. *JAMA Intern Med*. 2020 Apr 7. doi: 10.1001/jamainternmed.2020.1447.
41. Ioannidis JPA, Axford CA, Contopoulos-Ioannidis JP. Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environmental Research* 2020 (in press).
42. Williamson E, Walker AJ, Bhaskaran KJ, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked

1
2
3
4 electronic health records of 17 million adult NHS patients. medRxiv.
5
6 2020:2020.05.06.20092999.
7
8

- 9 43. Saltelli A, Bammer G, Bruno I, Charters E, Di Fiore M, Didier E, Nelson Espeland W,
10 Kay J, Lo Piano S, Mayo D, Pielke R Jr, Portaluri T, Porter TM, Puy A, Rafols I, Ravetz
11 JR, Reinert E, Sarewitz D, Stark PB, Stirling A, van der Sluijs J, Vineis P. Five ways to
12 ensure that models serve society: a manifesto. Nature. 2020;582:482-484.
13
14 44. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Whittaker C, Zhu
15 H, Berah T, Eaton JW, Monod M; Imperial College COVID-19 Response Team, Ghani
16 AC, Donnelly CA, Riley SM, Vollmer MAC, Ferguson NM, Okell LC, Bhatt S.
17 Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.
18 Nature. 2020 Jun 8. doi: 10.1038/s41586-020-2405-7.
19
20 45. Homburg S, Kuhbandner C. Comment on Flaxman et al. (2020, Nature,
21 <https://doi.org/10.1038/s41586-020-2405-7>): The illusory effects of non-pharmaceutical
22 interventions on COVID-19 in Europe. In: [https://s3-eu-west-1.amazonaws.com/pstorage-advance-](https://s3-eu-west-1.amazonaws.com/pstorage-advance-4589058904380/23182457/2020CommentFlaxmanPreprint.pdf)
23 [4589058904380/23182457/2020CommentFlaxmanPreprint.pdf](https://s3-eu-west-1.amazonaws.com/pstorage-advance-4589058904380/23182457/2020CommentFlaxmanPreprint.pdf), last accessed June 20,
24 2020
25
26 46. [https://www.healio.com/news/infectious-disease/20200506/covid19-will-set-fight-](https://www.healio.com/news/infectious-disease/20200506/covid19-will-set-fight-against-tb-back-at-least-5-years)
27 [against-tb-back-at-least-5-years](https://www.healio.com/news/infectious-disease/20200506/covid19-will-set-fight-against-tb-back-at-least-5-years), last accessed June 20, 2020.
28
29 47. <https://apps.who.int/iris/handle/10665/331590>, last accessed June 20, 2020.
30
31 48. US could see 100,000 new COVID-19 cases per day, Fauci says.
32
33 [https://www.statnews.com/2020/06/30/u-s-could-see-100000-new-covid-19-cases-per-](https://www.statnews.com/2020/06/30/u-s-could-see-100000-new-covid-19-cases-per-day-fauci-says/)
34 [day-fauci-says/](https://www.statnews.com/2020/06/30/u-s-could-see-100000-new-covid-19-cases-per-day-fauci-says/), last accessed July 5, 2020.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 49. Actual Coronavirus Infections Vastly Undercounted, C.D.C. Data Shows - The New
5
6
7 York Times, [https://www.nytimes.com/2020/06/27/health/coronavirus-antibodies-](https://www.nytimes.com/2020/06/27/health/coronavirus-antibodies-asymptomatic.html)
8
9 asymptomatic.html, last accessed July 5, 2020.
10
11
12 50. [https://cdn.ymaws.com/www.cste.org/resource/resmgr/2020ps/Interim-20-ID-](https://cdn.ymaws.com/www.cste.org/resource/resmgr/2020ps/Interim-20-ID-01_COVID-19.pdf)
13
14 [01_COVID-19.pdf](https://cdn.ymaws.com/www.cste.org/resource/resmgr/2020ps/Interim-20-ID-01_COVID-19.pdf), last accessed June 20, 2020
15
16
17 51. Roxby AC, Greninger AL, Hatfield KM, Lynch JB, Dellit TH, James A, Taylor J, Page
18
19 LC, Kimball A, Arons M, Schieve LA, Munanga A, Stone N, Jernigan JA, Reddy SC,
20
21 Lewis J, Cohen SA, Jerome KR, Duchin JS, Neme S. Detection of SARS-CoV-2 Among
22
23 residents and staff members of an independent and assisted living community for older
24
25 adults - Seattle, Washington, 2020. MMWR Morb Mortal Wkly Rep. 2020 Apr
26
27 10;69(14):416-418.
28
29
30
31 52. Treibel TA. Manisty C, Burton M, et al. Covid-19: PCR screening of asymptomatic
32
33 health-care workers at London hospital. Lancet 2020;395:1608-1610.
34
35
36 53. Sandri MT, Azzolini E, Torri V, Carloni S, Tedeschi M, Castoldi M, Mantovani A,
37
38 Rescigno M. IgG serology in health care and administrative staff populations from 7
39
40 hospital representative of different exposures to SARS-CoV-2 in Lombardy, Italy
41
42 . medRxiv 2020, doi: <https://doi.org/10.1101/2020.05.24.20111245>
43
44
45 54. Mansour M, Leven E, Muellers K, Stone K, Mendu DR, Wainberg A. Prevalence of
46
47 SARS-CoV-2 antibodies among healthcare workers at a tertiary academic hospital in
48
49 New York City. J Gen Intern Med. 2020 Jun 3 : 1–2. doi: 10.1007/s11606-020-05926-8
50
51
52 55. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, et al. Presymptomatic SARS-
53
54 CoV-2 infections and transmission in a skilled nursing facility. N Engl J Med.
55
56 2020;382:2081-2090.
57
58
59
60
61
62
63
64
65

- 1
2
3
4 56. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' Heel of
5
6 current strategies to control covid-19. N Engl J Med. 2020;382:2158-2160.
7
8
9 57. McIntosh K. Coronaviruses. UpToDate.
10
11 <https://www.uptodate.com/contents/coronaviruses>, last accessed June 19, 2020.
12
13
14 58. Grisin S. Covid-19: “Staggering number” of extra deaths in community is not explained
15
16 by covid-19. BMJ 2020; 369 doi: <https://doi.org/10.1136/bmj.m1931>.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 <u>their projections were alarming</u> . 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 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)	“In California alone, at least 1.2 million people over the age	“In our home state of California, for

<p>and https://medicalxpress.com/news/2020-04-opinion-hospitals-beds-non-covid-patients.html)</p>	<p>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.”</p>	<p>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.”</p>
---	--	--

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 (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 122.7 thousand instead of 350 thousand. 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</p>	<p>Within less than 4 weeks of this quote the number of</p>

<p>that somewhere between 800 to 1,000 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>daily deaths was much less than the 800-1000 quote (516 daily average for the week ending July 4). Then it increased again. The number of actual total deaths as of September will be added here when available.</p>
--	---

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 daily the number of deaths. They may vary because of reporting delays, changing definitions, data errors, and more reasons. 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, estimated 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 on 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	While some interventions in the broader package of lockdown measures are likely to have beneficial effects, assuming huge benefits is incongruent with the past (weak) evidence and should be avoided. Large benefits may be feasible from

	(influenza) pathogen that had ~100-fold higher infection fatality rate than SARS-CoV-2. Even thus, the impact on reduction on total deaths was of borderline significance and very small (10-20% relative risk reduction); conversely many models have assumed 25-fold reduction in deaths (e.g. from 510,000 deaths to 20,000 deaths in the Imperial College model) with adopted measures	precise, focused measures (e.g. early, intensive testing with through 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.
Lack of transparency	Many models used by policy makers were not disclosed as to their methods; 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 take more time unavoidably, 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; since 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 on the same table. Major pandemics happen rarely and what is needed are models which fuse information from a variety of sources. Information from data, from experts in the field, from past pandemics, need to be fused 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
---------------------	--	---

Table 4. Adverse consequences of what sounds like a sensible precautionary action: prepare hospitals to encounter an excess of hospitalization and ICU needs far beyond their standard capacity (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, ⁶ 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 ³³ 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, ⁵ 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 ⁵⁴
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

Table 5. Taleb's main statements and our responses

Forecasting single variables in fat tailed domains is in violation of both common sense and probability theory.	<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 $P(a < X < B)$.
Pandemics are extremely fat tailed.	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. ²⁷
Science is not about making single points predictions but understanding properties (which can <i>sometimes</i> be tested by single points).	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.
Risk management is concerned with tail properties and distribution of extrema, not averages or survival functions.	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 examining the NY daily death count data studied in ref. 12, these data are found to be characterized as stochastic rather than chaotic. ²⁸
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.	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 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. Consequent on this forecast a decision was made to send COVID19 patients to nursing homes, with tragic consequences.

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 COVID-19 becoming a notifiable disease,⁵⁰ despite being aware of the Italian experience⁴⁰ where almost all counted "COVID-19 deaths" had also 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 state⁵¹ 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 with draconian measures: nursing homes and hospitals. Over several weeks, I tried unsuccessfully to publish this in three medical journals and in 5 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 entirely the first half, stated it will appear next day, then said they were sorry. 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 "no infectious disease expert thinks this way" – paradoxically, I am trained and certified in infectious diseases.

45-53% of deaths in the US³⁸ (and as many or more in several European countries)³⁹ eventually were in nursing homes and related facilities and probably another large share were nosocomial, hospital-acquired 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 workers⁵²⁻⁵⁴ and in nursing homes.^{55,56}

Had we dealt with this coronavirus thinking what other widely-circulating coronaviruses do based on medical or infectious disease (not modeling) textbooks (=they cause mostly mild infections, but they can devastate specifically nursing homes and hospitals),^{20,21,57} 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.

FIGURE LEGENDS

Figure 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 over predicted enormously, and then it under predicts. For New Jersey, a neighboring state, the model started well but then it underpredicts, while for California it predicted a peak which never eventuated.

Figure 2: Performance of four data-driven models, IHME, YYG, UT, 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 hours 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%.

Figure 3: Performance of the same models examined in Figure 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.

Figure 4: Snapshot from <https://reichlab.io/covid19-forecast-hub/> (a very useful site that collates information and prediction from multiple forecasting models) as of 11.14am 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, a 4.5-fold difference, and the spectrum of 95% confidence intervals ranged from fewer than 100 deaths to over 16,000 deaths, almost a 200-fold difference.

Figure 5: Population age-risk categories and COVID-19 deaths per age-risk category. The illustration uses estimates for symptomatic case fatality rate of 0.05% in age 0-49, 0.2% in age 50-64, and 1.3% in age 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 age 0-49, 30% are asymptomatic in age 50-64, and 10% are asymptomatic in age 65 and over. Furthermore it assumes that among people in nursing homes and related facilities (~0.5% of the population in the USA), infection fatality rate is 26%, as per [55]. 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 over category. Based on available prognostic models (e.g. [42]) this prognostic classification should be readily attainable. As shown, <10% of the population is at high risk (shown with dense-colors and thus worth of special protection and more aggressive measures), and these people account for >90% of the potential deaths. >90% of the population could possibly continue with non-disruptive measures, since they account for only <10% of the total potential deaths.

Figure 1

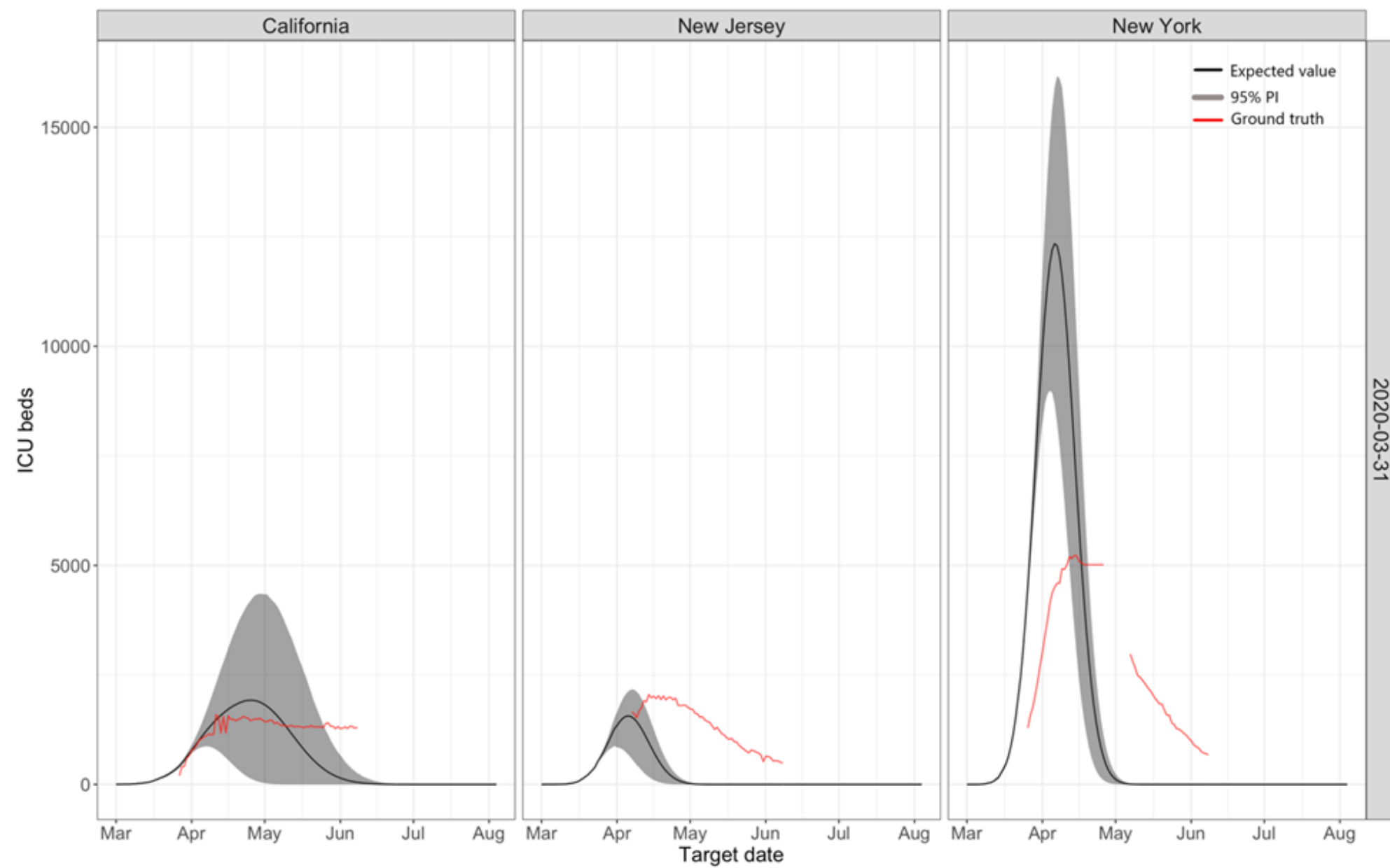


Figure 2

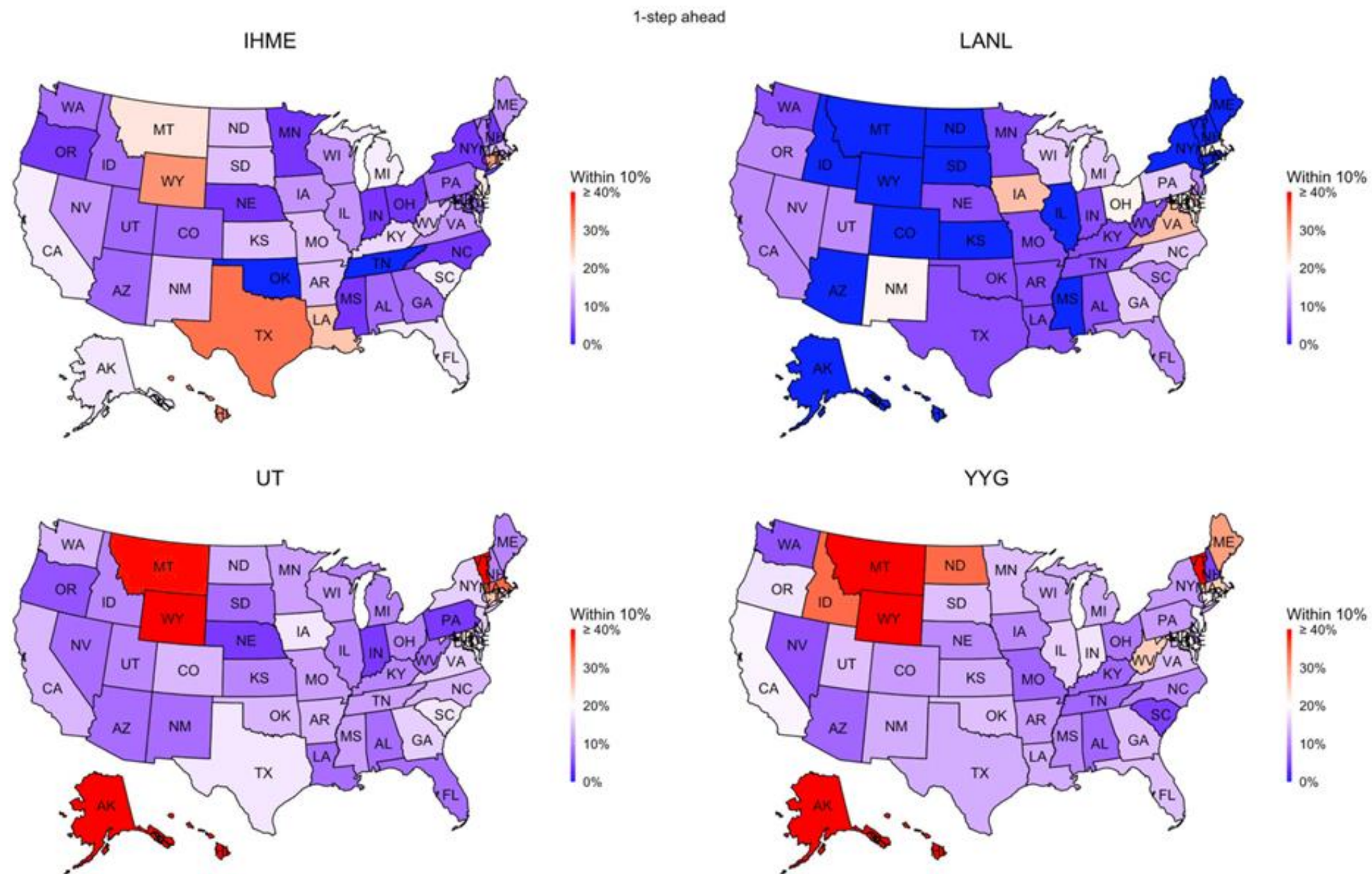


Figure 3

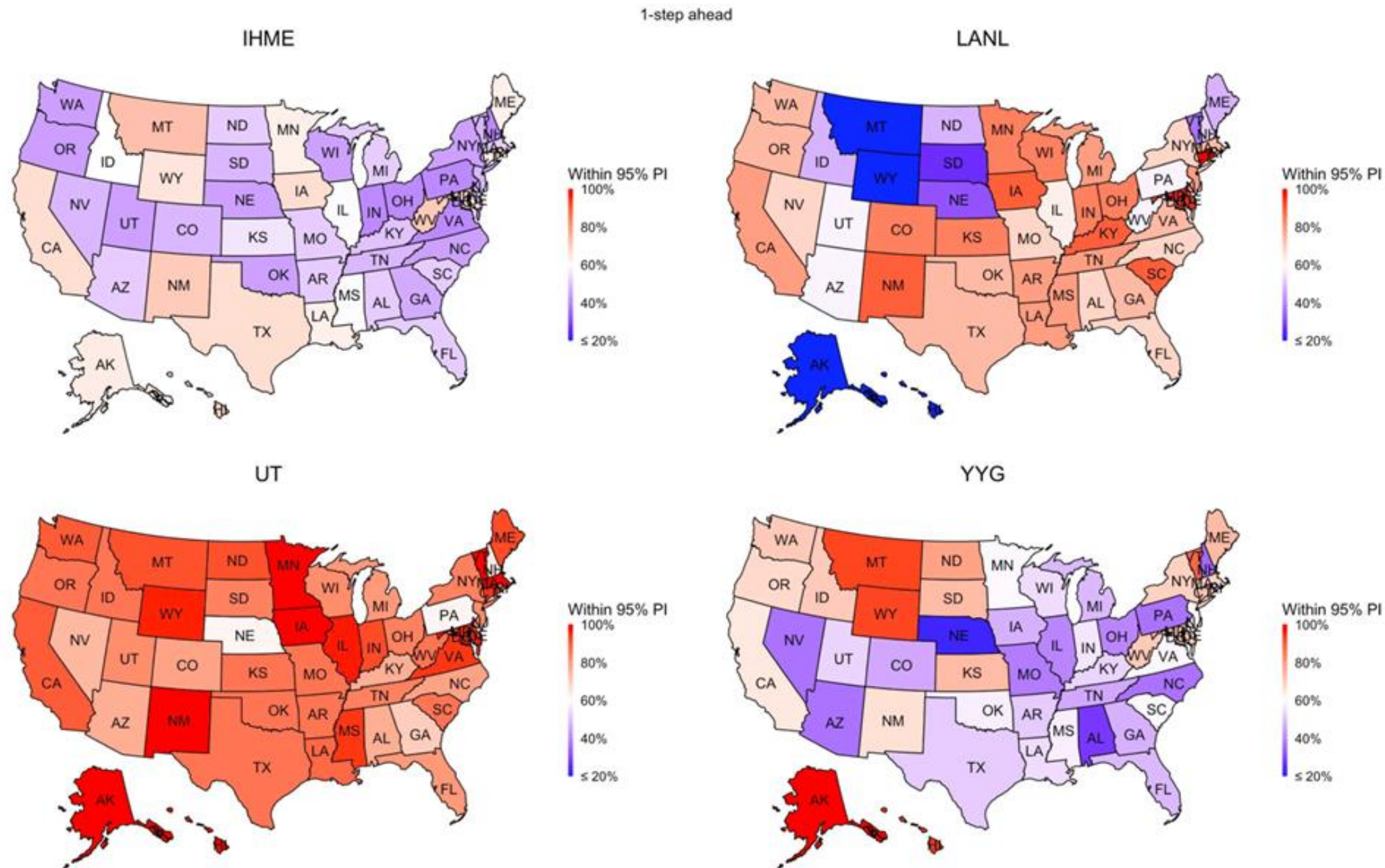


Figure 4

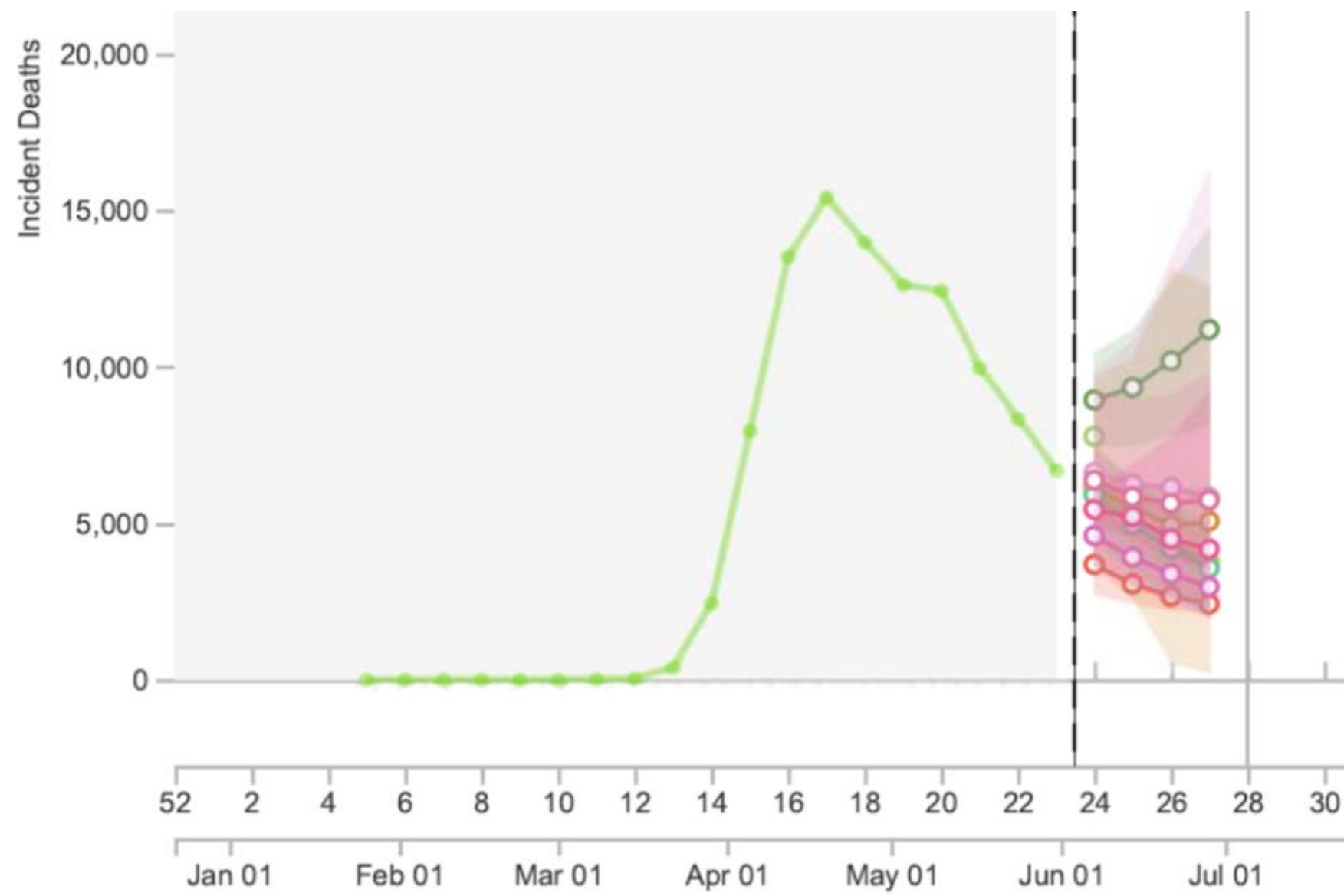


Figure 5

