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

COVID-19

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

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

Received: 14 June 2020 / Accepted: 21 July 2020 / Published online: 11 August 2020

Abstract

Forecasting models have been influential in shaping decision-making in the COVID-19 pandemic. However, there is concern that their predictions may have been misleading. Here, we dissect the predictions made by four models for the daily COVID-19 death counts between March 25 and June 5 in New York state, as well as the predictions of ICU bed utilisation made by the influential IHME model. We evaluated the accuracy of the point estimates and the accuracy of the uncertainty estimates of the model predictions. First, we compared the "ground truth" data sources on daily deaths against which these models were trained. Three different data sources were used by these models, and these had substantial differences in recorded daily death counts. Two additional data sources that we examined also provided different death counts per day. For accuracy of prediction, all models fared very poorly. Only 10.2% of the predictions fell within 10% of their training ground truth, irrespective of distance into the future. For accurate assessment of uncertainty, only one model matched relatively well the nominal 95% coverage, but that model did not start predictions until April 16, thus had no impact on early, major decisions. For ICU bed utilisation, the IHME model was highly inaccurate; the point estimates only started to match ground truth after the pandemic wave had started to wane. We conclude that trustworthy models require trustworthy input data to be trained upon. Moreover, models need to be subjected to prespecified real time performance tests, before their results are provided to policy makers and public health officials.

Keywords COVID-19 · Hospital resource utilisation · Model evaluation · Uncertainty quantification

✉ Sally Cripps
sally.cripps@sydney.edu.au

1 ARC Centre for Data Analytics for Resources and Environments, Sydney, Australia

2 School of Mathematics and Statistics, The University of Sydney, Sydney, Australia

3 Department of Statistics, Northwestern University, Chicago, USA

4 Department of Mathematical Sciences, University of Texas at El Paso, El Paso, USA

5 Stanford Prevention Research Center, Stanford, USA

6 Department of Medicine, Stanford University, Stanford, USA

7 Department of Epidemiology and Population Health, Stanford University, Stanford, USA

8 Department of Biomedical Data Sciences, Stanford University, Stanford, USA

9 Department of Statistics, Stanford University, Stanford, USA

10 Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, USA

Introduction

> I don't have a crystal ball. Everybody's entitled to their own opinion, but I don't operate here on opinion. I operate on facts and on data and on numbers and on projections. [12] New York Governor Andrew Cuomo - March 24, 2020
>
> Now, people can speculate. People can guess. I think next week, I think two weeks, I think a month, I'm out of that business because we all failed at that business. Right? All the early national experts. Here's my projection model. Here's my projection model. They were all wrong. They were all wrong. [7] New York Governor Andrew Cuomo - May 25, 2020

Forecasting has been very influential in the COVID-19 pandemic. Dealing with a new virus and with a lot of uncertainties surrounding its eventual impact, policy makers have widely used and depended upon predictions made by various

models. These predictions refer to critical issues such as the number of anticipated deaths with and without different interventions and the number of hospital beds, ICU beds, and ventilators that would be needed to deal with the surge of the epidemic waves. There is concern that while these models are useful, they can also be very misleading [13, 15, 16]. It is important to understand their performance and their limitations and to try to learn from their failures. This may help generate some better standards for the construction, validation, and use of these models.

In this article, we evaluate four models for predicting the daily death counts attributable to COVID-19 for the period March 25 to June 5 for the state of New York (NY), as well as one early model that predicted ICU bed utilisation in NY. The models evaluated are those constructed by the Institute of Health Metrics and Evaluations (IHME) [14], Youyang Gu (YYG) [11], the University of Texas at Austin (UT) [27], and the Los Alamos National Laboratory (LANL) [17]. These models were chosen because they provide daily death count predictions, as well as 95% prediction intervals (PIs) for each prediction. The IHME model began producing forecasts from March 25, the corresponding dates for YYG, UT and LANL are April 2, April 14 and April 16, respectively. We evaluate these models based on two criteria. The first criterion is the accuracy of the point estimates and the second criterion is the accuracy of the uncertainty estimates of those predictions. With regard to accuracy of prediction, we do not find a model that distinguishes itself from the pack. Most concerning, across models only 10.2% of the predictions fall within 10% of their training ground truth, irrespective of distance into the future. For accurate assessment of uncertainty, the LANL model had observed coverage most closely matching the nominal 95% coverage. Unfortunately, the LANL model did not commence predictions until April 16, approximately a month after the final US state declared a state of emergency and eleven days after the final US state entered lock-down, thus it played no role in the initial major decisions made by key policy makers in NY, as well as Washington DC. Regarding the prediction of ICU bed utilisation, the single model (IHME) was highly inaccurate and the point estimates only started to match ground truth by early May, after the pandemic wave had started to wane. Two major takeaways from this research are that

1. Serious thought and investment should be made in quality data collection when it comes to COVID-19 daily death data, as well as COVID-19 resource utilisation.
2. Models need to be subjected to real time performance tests, before their results are provided to policy makers and public health officials. In this paper, we provide examples of such tests, but irrespective of which tests are adopted, they need to be specified in advance, as one would do in a well-run clinical trial.

The data

In order to evaluate the models, it is necessary to define the actual ground truth number of daily deaths. This task is more problematic than it would first appear, as there is no one source of ground truth. The models YYG and LANL use the raw daily death counts in NY reported by the Johns Hopkins University (JHURD) [4] for training, IHME uses daily deaths reported by the New York Times (NYT) [24], while UT uses NYT data until May 5 to train their model before switching to another version of the JHU data which is known as the JHU time series (JHUTS) data [5]. The JHUTS data is an update of the JHURD to correct for reporting errors. Many modellers (e.g. [11]) view the JHURD data as the gold standard, though [17] raise concerns with the JHURD data, as well.

Figure 1 presents the ground truth data for the state of NY as reported by JHURD (red), JHUTS (dark blue), NYT (green), as well as two additional sources: CovidTracking [23] (black) and USAFacts [26] (light blue), from March 15 to June 5. The point of this figure is to demonstrate that these ground truths can vary substantially from each other and have features which are artefacts of the way in which deaths are reported. Of particular note is the early lag between JHURD and JHUTS, as well as the more smoothed process presented by the NYT data. It is noted that the NYT data are the confirmed COVID-19 cases from January 22 to May 6, while from May 7 on-wards, the NYT data include the confirmed and probable cases using criteria that were developed by local and states government [25]. Both JHURD and JHUTS include confirmed and probable cases. In all sources, the actual ground truth number of daily deaths is calculated by taking the difference of cumulative deaths.

Figure 1 also shows evidence of large swings in the number of reported daily deaths, again probably due to lags or corrections in reporting. Indeed, the JHUTS data show a negative value for the number of deaths in NY on April 19. This negative value is, in turn, a result of updates to the JHURD to correct reporting errors [6]. This value is clearly incorrect. However, without information regarding the details of cases at the individual level, it is not possible to correct these data. Any attempt to smooth the data raises the question of how the choice of smoothing technique may affect any conclusions drawn from the data and [18] raise numerous concerns in this regard. In summary, coding, counting and reporting COVID-19 deaths is highly complex and is beyond the scope of this paper (see e.g. [19]). Accordingly, we have chosen to work with the data provided by these sources and will evaluate each model according to the data the developers have chosen to use for training purposes, as well as to all three ground truths, namely, NYT, JHURD and JHUTS.

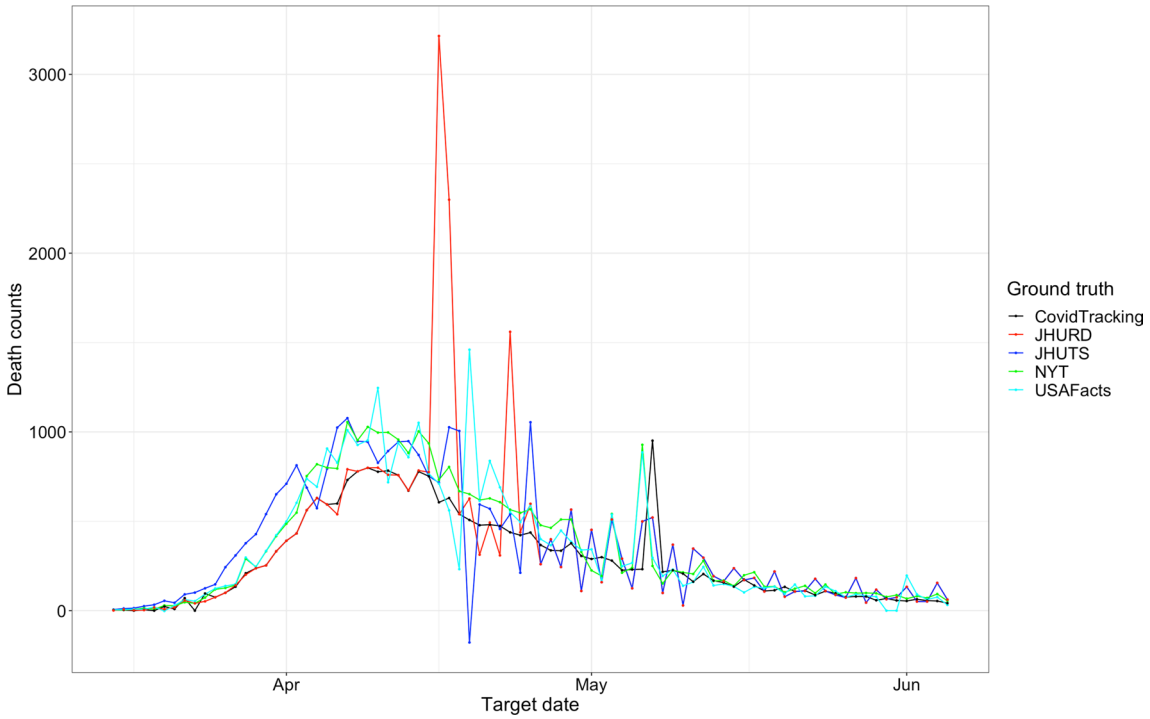


Fig. 1 A comparison of the daily death counts ground truth from CovidTracking (black), JHURD (red), JHUTS (dark blue), NYT (green) and USAFacts (light blue) for the period March 15–June 5 for NY

Accuracy of the point etimates

Figure 2 shows the actual data used to develop each of the models, as well as the time series of forecasts made by each of the models. One can see the spike in the number of deaths reported by the NYT in early May following the inclusion of probable cases. Figure 2 displays only the point estimates of the forecasts. The time series of forecasts are colour-coded such that the earliest/latest forecasts are at the red/blue end of the colour spectrum. For example, the deepest red curves are the time series of forecasts made in late March, the yellow curves are time series of forecasts made in early April and so on, until the violet curves which represent the most recent time series of forecasts.

To evaluate each of the forecast time series in Fig. 2, we computed two metrics for each forecast. These metrics are the mean absolute percentage error and the maximum absolute percentage error, whereby the percentage error is computed from $(ground\ truth—predicted\ value)/(ground\ truth) \times 100\%$. For the mean absolute percentage error, the percentage of discrepancy between the given model's prediction and the ground truth was computed for each model for each day for each time series. This information was then averaged over the entire duration of the forecast for a particular time series. The maximum absolute percentage error was computed by taking the maximum of the absolute percentage errors for each forecast and for each model. For example, the first forecast made by IHME was on March 25 and we compare the forecast time series for the period from March 25 until June 5 with the ground truth time series over that same period by calculating the two metrics discussed above. We then repeat the process for each date a forecast time series was issued, and for each of the models. To make a fair comparison on dates where the forecast time series was made by at least two of the models, we truncate the forecast time series at the last prediction date of the shortest time series. These values are plotted over time in Fig. 3 for each version of the ground truth.

In Fig. 3, we see that while some models may perform better or worse over subsets of the time frame of interest, no one model clearly dominates throughout with respect to either of the metrics for any version of the ground truth data.

Accuracy of uncertainty quantification

We now turn to the subject of uncertainty quantification of these models. Each of the models provides estimates of uncertainty, where the IHME, YYG and LANL forecasts give 95% PIs, while the UT provides 90% PIs for predictions made prior to the forecast date of May 16. To translate these 90% PIs to 95% PIs for the UT model, we take the log of the prediction and the 90% PIs; calculate the difference between the log of the prediction and the log of the 90% PI limits and

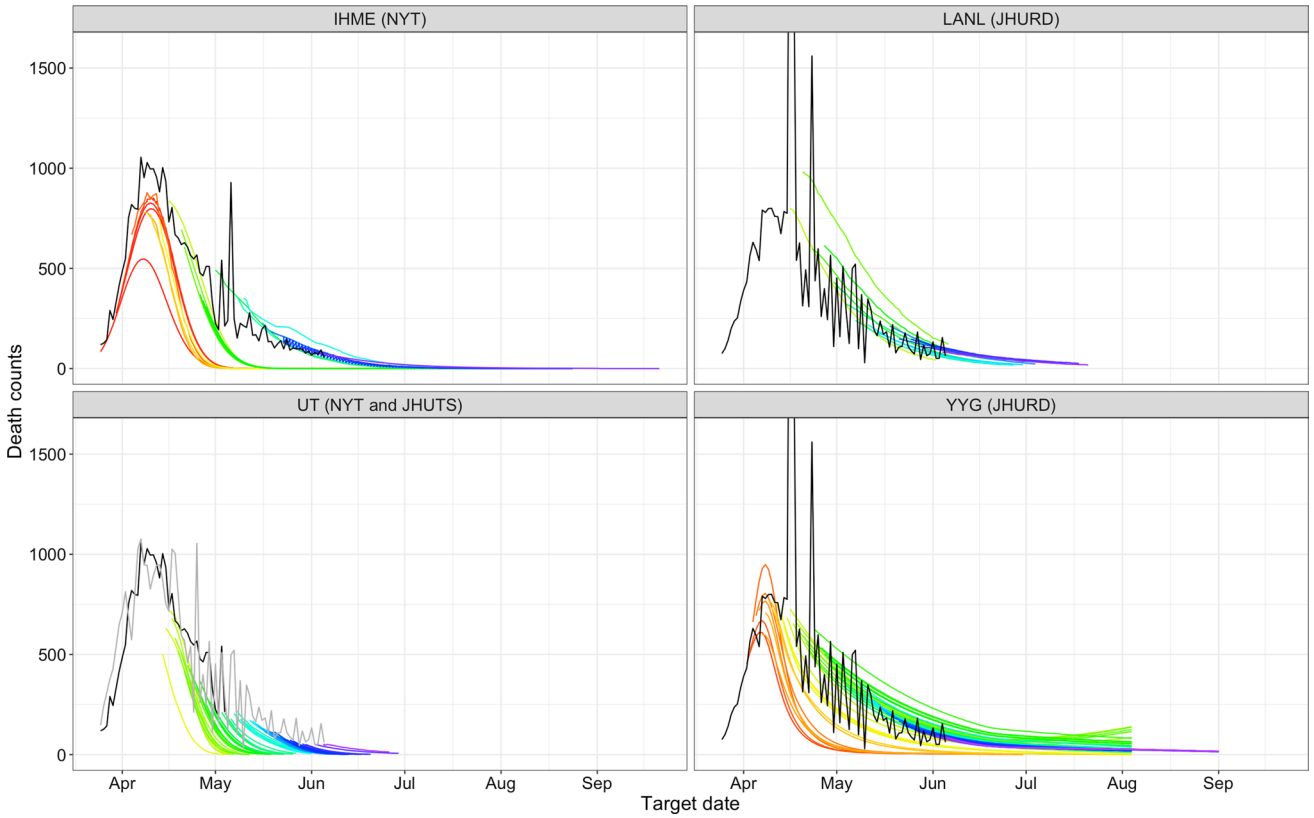


Fig. 2 The forecast time series made by each model, along with the ground truth (black) used to train each model. The UT model uses the NYT data (black) until May 5 before switching to the JHUTS data (grey), whereby the negative value for the daily deaths on April 19 (see Fig. 1) is removed before the model is trained

multiply this difference by a factor of 1.96/1.64. We recompute the 95% PIs on the log scale before transforming them back to the original scale.

Figure 4 presents plots of the 95% PIs for various predictions made by the models and the training ground truth. We follow [18] and define a *k-step-ahead* prediction and PI for a particular date, to be the prediction and accompanying PI made k days in advance of that date. For example, for June 3, a 1-step-ahead prediction and PI are the prediction and accompanying interval for June 3 made on the forecast date of June 2, while the 2-step-ahead prediction and PI for June 3 would be made on the forecast date of June 1, etc. The columns of Fig. 4 relate to the number of step-ahead predictions ranging from 1 to 7 in panel 4a and from 8 to 14 in panel 4b, while the rows of Fig. 4 correspond to the different models.

Figure 4 shows a number of interesting features. First, as documented in [18], the IHME model undergoes a number of dramatic changes in the calculation of the prediction and the corresponding PIs. The original IHME model underestimates uncertainty and 45.7% of the predictions (over 1- to 14-step-ahead predictions) made over the period March 24 to March 31 are outside the 95% PIs. The IHME model was revised on April 2 and made no predictions on April 1 and April 2. In the revised model, for forecasts from of April 3 to May 3 the uncertainty bounds are enlarged, and most predictions (74.0%) are within the 95% PIs, which is not surprising given the PIs are in the order of 300 to 2000 daily deaths. Yet, even with this major revision, the claimed nominal coverage of 95% well exceeds the actual coverage. On May 4, the IHME model undergoes another major revision, and the uncertainty is again dramatically reduced with the result that 47.4% of the actual daily deaths fall outside the 95% PIs—well beyond the claimed 5% nominal value. It is concerning, nevertheless, that the uncertainty estimates of the IHME model seem to improve with the forecast horizon, so that for the original model and latest IHME model update, more observed values fall within the 95% PIs for the 7-step-ahead prediction than for the 1-step-ahead prediction.

Second, Fig. 4 shows that the YYG model does not perform well in terms of uncertainty quantification, as there are many more actual deaths lying outside the 95% PIs than would be expected. Taken across the entire time period, the proportion of actual deaths lying outside the 95% PIs is 31.1%. We do, however, note that this percentage improves over time. From the forecast date of May 1 on-wards, the fraction of actual deaths lying outside the

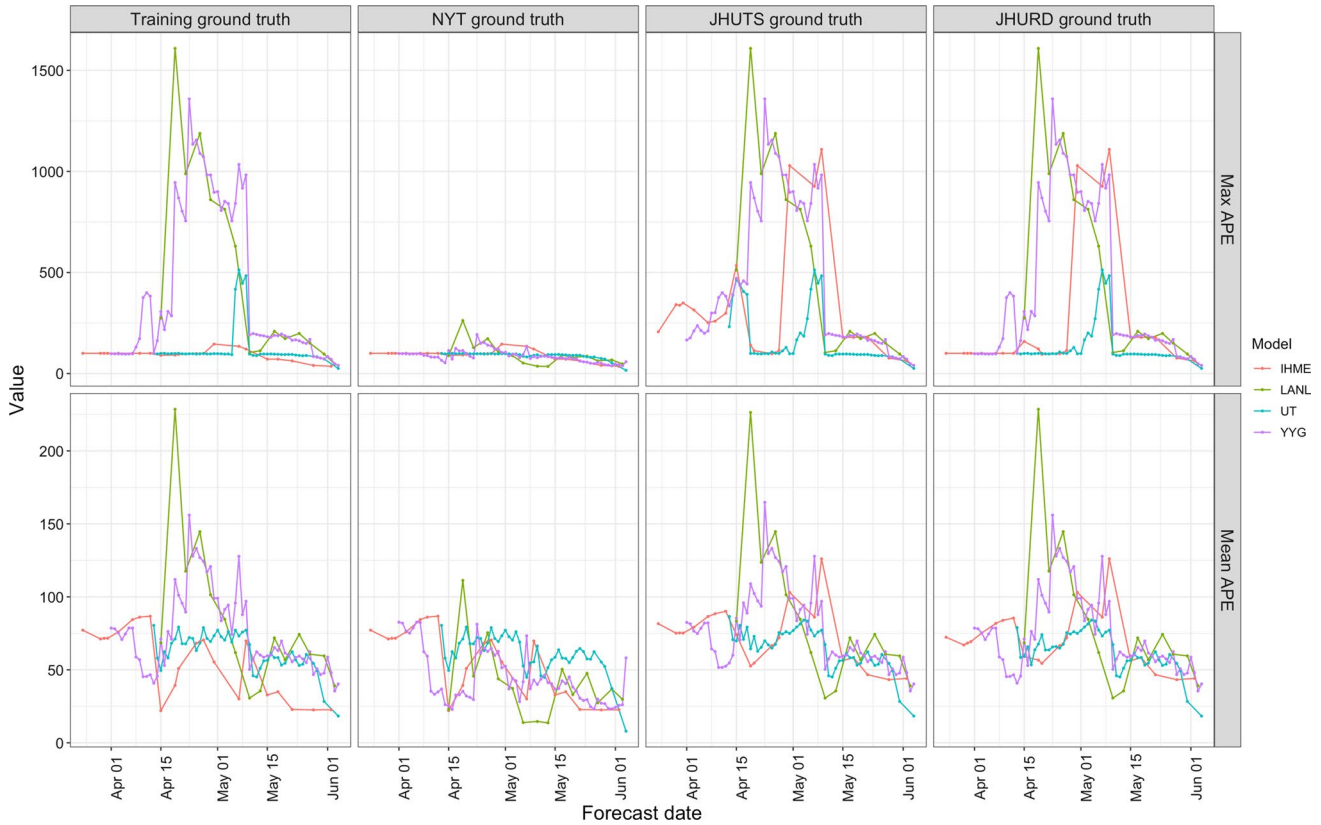


Fig. 3 Discrepancies between each model and the ground truth, as measured by the maximum absolute percentage error (top) and the mean absolute percentage error (bottom), for each version of the ground truth

95% PIs ranges from 28.6% for the 1-step-ahead prediction to 13.6% for the 14-step-ahead prediction, in comparison to 44.8% for the 1-step-ahead prediction to 48.3% for the 14-step-ahead prediction prior to this date.

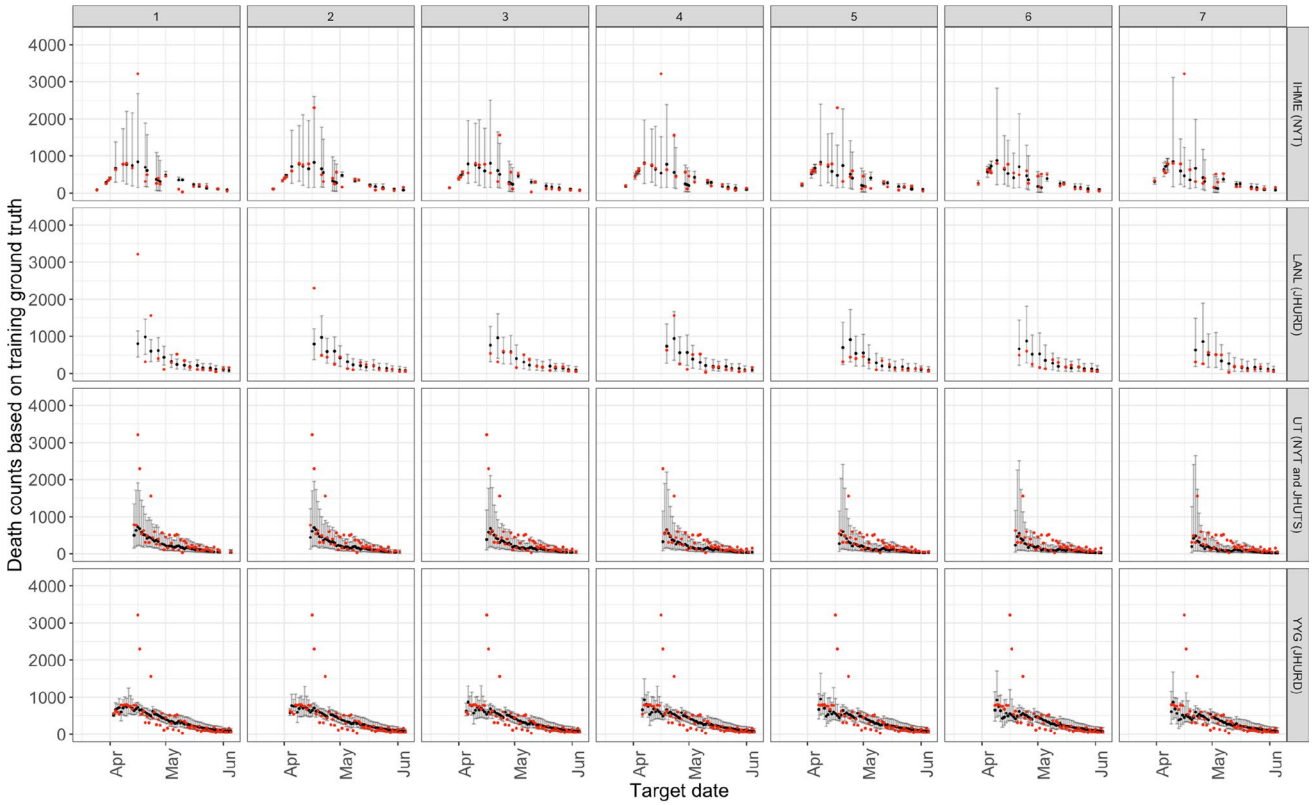
Similarly, regarding the UT and LANL forecasts, neither has observed coverage consistently matching the 95% nominal coverage as shown in Fig. 5 (first plot in the top panel). For the UT model, the fraction of actual deaths lying outside the 95% PIs ranges from 14.6% for the 1-step-ahead prediction to 67.5% for the 14-step-ahead. The corresponding figures for the LANL are 40.0% for the 1-step-ahead prediction to 9.1% for the 14-step-ahead.

The second, third and fourth plots in the top panel (one for each version of ground truth) in Fig. 5 present this information from a slightly different perspective. In particular, from this point of view, for the 5-14 step-ahead predictions, LANL had the best observed coverage compared to the nominal 95% level, with short-term predictions tending to overestimate the ground truth. The PIs for the UT model are seen to deteriorate for predictions out into the future, with a tendency of the model to underestimate the daily number of deaths. The remaining two models, YYG and IHME, tended to provide daily death prediction PIs that systematically miss the nominal 95% coverage level, irrespective of distance into the future.

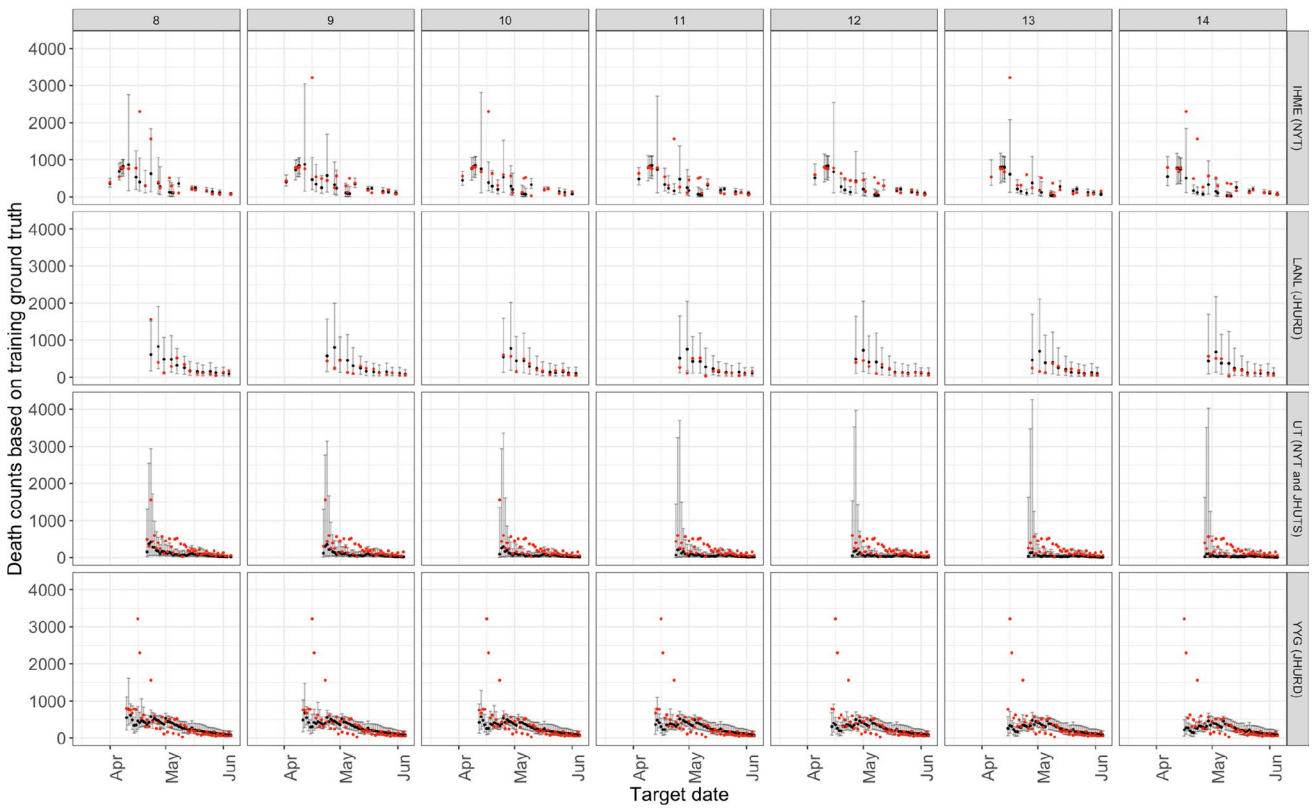
Examining the last panel in Fig. 5, with the exceptions of IHME evaluated on JHURD and LANL evaluated on the NYT ground truth, no model had more than 30% of daily death predictions falling within 10% of the ground truth, with the UT having virtually no predictions within a 10% bound of the ground truth, out into the future. When evaluated on their training ground truth, only 10.2% of the predictions fall within 10% of their training ground truth, irrespective of distance into the future.

Prediction of ICU bed utilisation

We now turn our attention to ICU bed utilisation in New York State. The only model that provides early daily predictions and PIs for NY ICU bed usage is the IHME model and we limit our attention to this model. The IHME model was very influential in early decision-making at the highest levels of the United States government, in regard to the allocation of resources for ICU bed usage, having been mentioned at



(a) 1- to 7-step-ahead predictions.



(b) 8- to 14-step-ahead prediction

Fig. 4 Different step-ahead predictions (black dots) by each model and their 95% PIs (gray bars), along with the ground truth (red dots) used to train each model

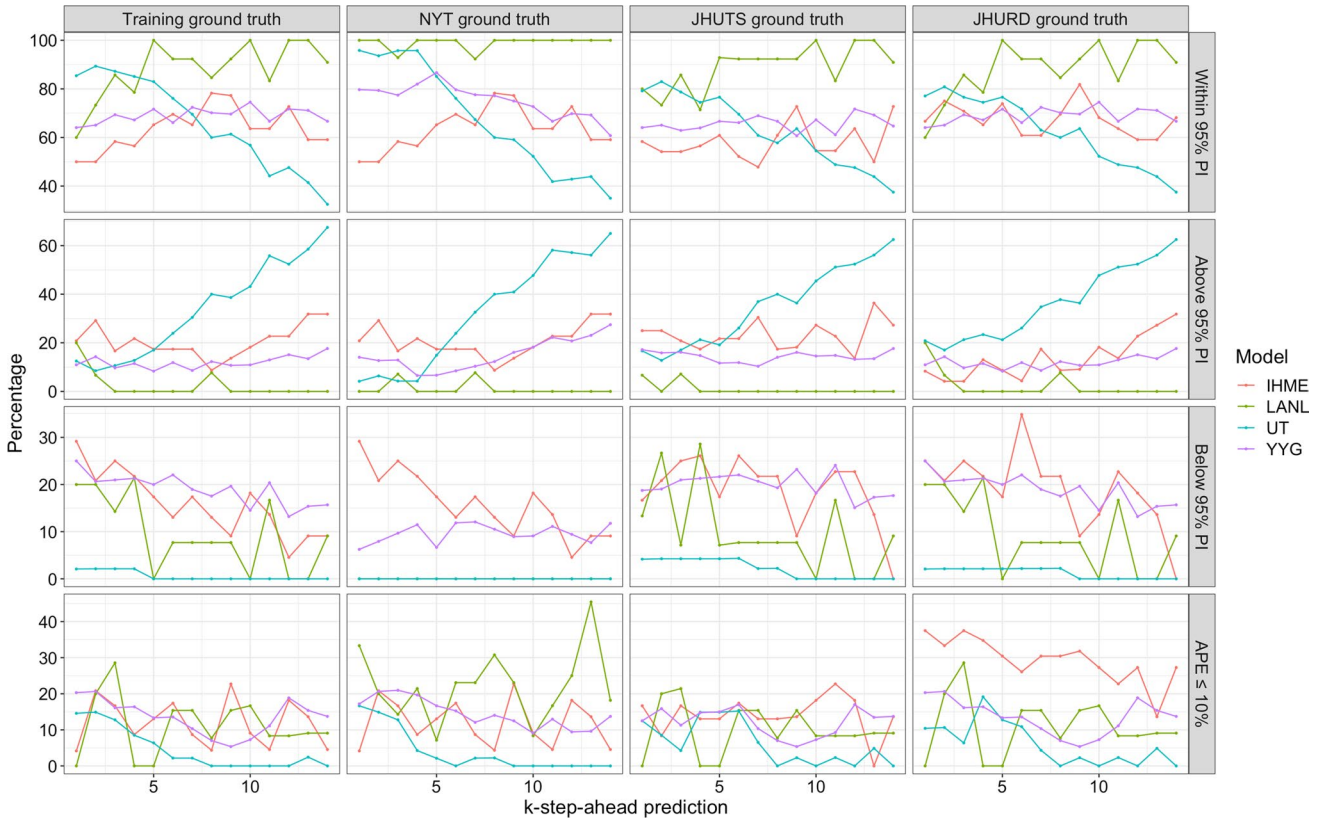


Fig. 5 Percentage of the number of daily deaths within, above and below the *k*-step-ahead 95% PIs. The last panel shows the percentage of *k*-step-ahead predictions which fall within 10% of the ground truth

White House Press conferences, including March 31, 2020 [28].

Figure 6 presents the IHME estimates (black) and 95% PIs (grey) for ICU bed usage in NY, together with the ground truth (red) and the maximum ICU capacity (blue; inclusive of non-COVID-19 ICU beds) obtained from THE CITY [22]. Each subplot in the figure corresponds to a day for which a prediction was made. For example, the first subplot is for the prediction made by IHME on March 25, the second for the prediction made on March 29, and so on until the last prediction made on June 5. The prediction intervals start at the date for which the prediction was made, and thus the gray shaded area, which represents the PIs, shifts to the right for subsequent subplots.

Figure 6 shows that the early forecasts of ICU bed utilisation were highly inaccurate–the prediction intervals for ICU beds made on March 25 through March 31 for the ICU bed usage on April 1 did not contain the actual value despite the width of these PIs being in the order of 5000 to 15,000. Over this period, the model seriously over-predicted the ICU bed usage. However, by the third week in April through early June, the point predictions of the IHME model systematically *underestimated* ICU bed utilisation. In fact, in late April, the model predicts zero bed usage by mid-May.

Conclusion

In a major crisis like COVID-19, policy makers and public health officials need to operate on "facts, data and numbers", but this can be difficult when these facts, data and numbers are highly error-prone. In the case of daily COVID-19 deaths in New York, there was serious disagreement even between sources regarding the ground truth for the number of deaths. A key take-away from our analysis is that *serious thought and investment must be made in quality data collection when it comes to COVID-19 daily death data, as well as COVID-19 resource utilisation* Clinical trial methodology [10] for data quality control must be brought to bear, especially when the consequences of policy decisions can so dramatically impact the lives of millions of people.

Early on, Dr. Anthony Fauci, NIAID Director, stated that [3]: "As I've told you on the show, models are really only as good as the assumptions that you put into the model. But when you start to see real data, you can modify that model..." An open question raised by this thoughtful comment is how can one expect quality predictions, when the data are suspect? How does one modify the model in light of the data, if the data are faulty? Would the course of action of policy leaders have differed, had it been known that there

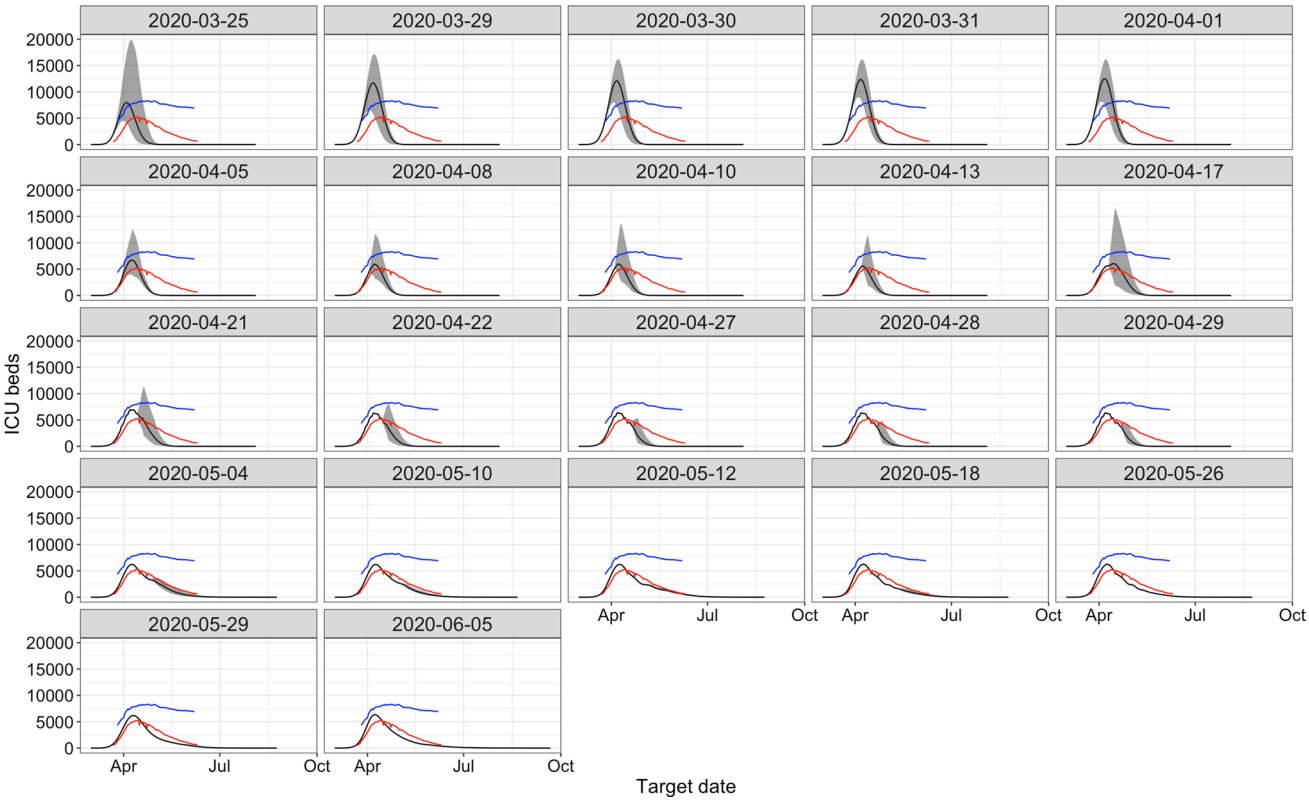


Fig. 6 Predicted ICU bed usage (black) and its 95% PIs (grey shaded area) in NY for each reporting date, along with the ground truth (red) and the maximum ICU capacity inclusive of non-COVID-19 ICU beds (blue) obtained from THE CITY

would not be clear agreement on what represented ground truth for even a hard endpoint such as death? Clearly, if the data are suspect, projections may also be sub-optimal.

However, putting the issue of data quality aside, our analysis shows that models tended to have very poor performance both in terms of accuracy as well as in terms of capturing uncertainty. To be fair, pandemics are, thankfully, rare events and predicting outcomes in the early stages is very difficult, as so much is unknown. Rosen [20] quotes Dr. Alain Labrique

> With a new virus, and any type of infectious disease that we have never encountered before, there are many unknowns," said Dr. Alain Labrique, an associate professor at The Johns Hopkins Bloomberg School of Public Health and one of the nation's most renowned epidemiologists." And so the big challenge for us is to focus on telling the public the truth about what we know, and to explain the uncertainties around what we don't know.

In this regard, the LANL model was the only model that was found to approach the 95% nominal coverage, but unfortunately this model was unavailable at the time Governor Cuomo needed to make major policy decisions in late March 2020. Model predictions for daily deaths tended to have smaller errors over time, but this is not reassuring, because predictions are extremely critical in the early phase of an epidemic wave.

The importance of accurate early predictions applies even more to predictions for bed utilisation, where wrong expectations can lead to wrong decisions. For example, a major mistake in New York was the decision to send COVID-19 patients to nursing homes. Based on a March 25 directive, over 4500 COVID-19 patients were discharged from hospitals to nursing homes [8], specifically because it was anticipated that regular hospital beds would be urgently needed and hospitals would be overrun by COVID-19 patients. Nursing homes are full of highly vulnerable people and outbreaks in nursing homes [1] resulted in high fatalities. In New York alone, over 5800 deaths occurred in nursing homes [8]. Eventually this was a sizeable fraction of the COVID-19 death burden, and importantly, it might have been avoidable to a large extent. Overestimates of anticipated bed requirements could also have affected hospital utilisation for other serious conditions with adverse consequences for the outcomes of patients suffering from these conditions [9, 21]. While

preparedness is important and beneficial, making preparations with vastly erroneous expectations can create major harm.

Our second key take-away from this evaluation:the need for real time evaluation of prediction models. Going forward there needs to be industry standards as to how models are to be evaluated and calibrated in real time, especially in the rapidly evolving settings of a pandemic. Quoting Dr. B. Jewell: "This appearance of certainty is seductive when the world is desperate to know what lies ahead" [2]. Unfortunately, in retrospect, COVID-19 anxiety can turn to COVID-19 disillusionment when the decisions made by policy makers are dictated by suspect low quality data and consequently by poorly performing models. One solution would be to compare predictions of models against emerging reality on a daily basis using prespecified metrics such as those analysed here. Models that are consistently poorly performing should carry less weight in shaping policy considerations. Models may be revised in the process, trying to improve performance. However, improvement of performance against retrospective data offers no guarantee for continued improvement in future predictions. Failed and recast models should not be given much weight in decision making until they have achieved a prospective track record that can instil some trust for their accuracy. Even then, real time evaluation should continue, since a model that performed well for a given period of time may fail to keep up under new circumstances.

References

1. Abrams HR, Loomer L, Gandhi A, Grabowski DC. Characteristics of US nursing homes with COVID-19 cases. J Am Geriatr Soc. 2020;. https://doi.org/10.1111/jgs.16661.
2. Begley S. Influential COVID-19 model uses flawed methods and shouldn't guide U.S. policies, critics say, April 17, 2020. https://www.statnews.com/2020/04/17/influential-covid-19-model-uses-flawed-methods-shouldnt-guide-policies-critics-say/. Accessed 7 June 2020.
3. Bump P. The lesson of revised death toll estimates shouldn't be that distancing was an overreaction, The Washington Post, April 9, 2020. https://www.washingtonpost.com/politics/2020/04/09/lesson-revised-death-toll-estimates-shouldnt-be-that-distancing-was-an-overreaction/. Accessed 7 June 2020.
4. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports_us. Accessed 7 June 2020.
5. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_US.csv. Accessed 7 June 2020.
6. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/README.md. Accessed 7 June 2020.
7. Cohen S. "We All Failed"—The real reason behind NY Governor Andrew Cuomo's surprising confession, Forbes, May 26, 2020. https://www.forbes.com/sites/sethcohen/2020/05/26/we-all-failed--the-real-reason-behind-ny-governor-andrew-cuomos-surprising-confession/#67f238f66fa5. Accessed 7 June 2020.
8. Condon B, Peltz J, Mustian J. AP count: over 4,500 virus patients sent to NY nursing homes, May 22, 2020. https://apnews.com/5ebc0ad45b73a899efa81f098330204c. Accessed 7 June 2020.
9. De Filippo O, D'Ascenzo F, Angelini F, Bocchino PP, Conrotto F, Saglietto A, Secco GG, Campo G, Gallone G, Verardi R, et al. Reduced rate of hospital admissions for ACS during COVID-19 outbreak in Northern Italy. N Engl J Med. 2020;. https://doi.org/10.1056/NEJMc2009166.
10. Friedman L, Furberg C, DeMets D, Reboussin D, Granger C. Fundamentals of clinical trials. 5th ed. Berin: Springer; 2015.
11. Gu Y. https://covid19-projections.com. Accessed 7 June 2020.
12. Herbert G. Cuomo refutes Trump, insists NY needs up to 40,000 ventilators: 'I operate on facts', Post-Standard, March 27, 2020. https://www.syracuse.com/coronavirus/2020/03/cuomo-refutes-trump-insists-ny-needs-up-to-40000-ventilators-i-operate-on-facts.html. Accessed 7 June 2020.
13. Holmdahl I, Buckee C. Wrong but useful—what COVID-19 epidemiologic models can and cannot tell us. N Engl J Med. 2020;. https://doi.org/10.1056/NEJMp2016822.
14. IHME COVID-19 health service utilization forecasting team and C. J. Murray. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. medRxiv, 2020. https://doi.org/10.1101/2020.04.21.20074732.
15. Ioannidis JP. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. Eur J Clin Investig. 2020;50(4):e13222. https://doi.org/10.1111/eci.13222.
16. Jewell NP, Lewnard JA, Jewell BL. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. JAMA. 2020;323(19):1893–4. https://doi.org/10.1001/jama.2020.6585.
17. LANL COVID-19 Team. https://covid-19.bsvgateway.org. Accessed 7 June 2020.
18. 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. arXiv:2004.04734v4, 2020.
19. Pappas S. How COVID-19 deaths are counted, Scientific American, May 19, 2020. https://www.scientificamerican.com/article/how-covid-19-deaths-are-counted1/. Accessed 7 June 2020.
20. Rosen J. Explaining Cuomo's lament: why coronavirus data models were 'all wrong'. https://wjla.com/news/nation-world/explaining-cuomos-lament-why-coronavirus-data-models-were-all-wrong. Accessed 7 June 2020.
21. Sud A, Jones ME, Broggio J, Loveday C, Torr B, Garrett A, Nicol DL, Jhanji S, Boyce SA, Ward P, et al. Collateral damage: the impact on cancer outcomes of the COVID-19 pandemic. 2020. https://doi.org/10.2139/ssrn.3582775.
22. THE CITY. https://thecity.nyc/. Accessed 7 June 2020.
23. The COVID Tracking Project. https://github.com/COVID19Tracking/covid-tracking-data/raw/master/data/states_daily_4pm_et.csv. Accessed 7 June 2020.
24. The New York Times. https://github.com/nytimes/covid-19-data. Accessed 7 June 2020.
25. The New York Times. https://github.com/nytimes/covid-19-data/blob/master/PROBABLE-CASES-NOTE.md. Accessed 7 June 2020.
26. USAFacts. https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_deaths_usafacts.csv. Accessed 7 June 2020.

27. UT Austin COVID-19 Modeling Consortium. https://covid19-projections.com. Accessed 7 June 2020.
28. White House press briefing on U.S. coronavirus response, March 31, 2020. https://www.youtube.com/watch?v=H7CoCiH2F7U&feature=youtu.be. Accessed 7 June 2020.