



Forecasting efforts from prior epidemics and COVID-19 predictions

Pranay Nadella¹ · Akshay Swaminathan² · S. V. Subramanian^{3,4} 

Received: 30 May 2020 / Accepted: 1 July 2020 / Published online: 17 July 2020
© Springer Nature B.V. 2020

Abstract

Since the onset of the COVID-19 pandemic, countless disease prediction models have emerged, shaping the focus of news media, policymakers, and broader society. We reviewed the accuracy of forecasts made during prior twenty-first century epidemics, namely SARS, H1N1, and Ebola. We found that while disease prediction models were relatively nascent as a research focus during SARS and H1N1, for Ebola, numerous such forecasts were published. We found that forecasts of deaths for Ebola were often far from the eventual reality, with a strong tendency to over predict. Given the societal prominence of these models, it is crucial that their uncertainty be communicated. Otherwise, we will be unaware if we are being falsely lulled into complacency or unjustifiably shocked into action.

Keywords Forecasting · Predictions · Pandemics · Ebola · COVID-19

Predictions of future cases, hospitalizations, and deaths have dominated the public discourse around COVID-19. Experts forecasted 20 to 60% of the world becoming infected and up to 2.2 million American deaths if the pandemic carries on unmitigated [1, 2]. Similarly, researchers predicted 510,000 British deaths and at least 300 million Indian cases [2, 3]. Many of these forecasts may not come true, so how can we use them effectively to inform policy?

Similar prediction efforts undertaken by researchers during past epidemics can lend clarity to this question. For example, prominent experts forecasted up to 200 million

and 50,000 global deaths for H5N1 and mad cow disease respectively [4, 5]. However, these were drastic overpredictions, as only 455 and 177 deaths ensued [6, 7].

To systematically evaluate the success of prior forecasting models, we reviewed predictions from three twenty-first century epidemics: the 2002–2004 Severe acute respiratory syndrome (SARS) outbreak, the 2009 H1N1 influenza pandemic, and the 2014 Ebola virus disease outbreak. We found that during the SARS and H1N1 outbreaks, only a few studies attempted to predict future cases and were ultimately unsuccessful. During the Ebola epidemic, the number of forecasting studies increased dramatically, and most overestimated—quite substantially—the true number of cases and deaths.

We identified studies that forecasted cases or deaths for each epidemic by employing a broad PubMed search strategy with the terms “estimate”, “model”, “forecast”, “predict”, “transmission”, and “intervention.” We only included studies that made predictions while the outbreak was occurring. Because not all forecasts may have been published in peer-reviewed literature, we applied this search strategy to news articles from major media outlets as well. For our analysis of the Ebola epidemic, we utilized the references from a prior Ebola review [8]. Of the reviewed studies, only three predicted deaths. For the remaining studies that predicted cases, we extrapolated deaths by multiplying cases with the studies’ estimates of case fatality rates (CFR). For studies that did not estimate CFR, we applied the average Ebola

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10654-020-00661-0>) contains supplementary material, which is available to authorized users.

✉ S. V. Subramanian
svsubram@hsph.harvard.edu

Pranay Nadella
pranay.nadella@pennmedicine.upenn.edu

Akshay Swaminathan
akshay325@gmail.com

- ¹ Center for Global Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
- ² Quantitative Sciences, Flatiron Health, New York, NY, USA
- ³ Center for Population and Development Studies, Harvard University, Cambridge, MA, USA
- ⁴ Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA, USA

CFR of 50%, as reported by the World Health Organization [9]. Finally, we numerically compared the studies' predicted deaths to eventual true deaths to assess prediction accuracy.

For SARS, only one prediction was found, which vastly overpredicted the number of cases in Canada by June 2003 (Predicted (P): 4432, Actual (A): 188) [10, 11]. During H1N1, there were only two studies that predicted future case counts and both significantly underpredicted the number of cases in the U.S. by June 2009 (P: 2000-2500, A: 100,000) [12, 13].

There were 17 studies (reporting 35 predictions) that forecasted the Ebola epidemic. Of the 35 predictions, 71% ($n=25$) overpredicted and 29% ($n=10$) underpredicted the number of deaths. These mispredictions varied from projecting 89% ($n=2256$) less than the actual deaths in Guinea to 9495% ($n=456,690$) more than the actual deaths in Liberia [14]. Only 37% ($n=13$) of the predictions were within the range of 50% greater or less than the actual number of deaths (Fig. 1). Additionally, several predictions were made assuming best case (all interventions implemented) or worst case (no interventions implemented) scenarios. Of the 12 predictions that assumed a worst-case scenario, 92% ($n=11$) overpredicted and 8% ($n=1$) underpredicted. Of the 7 predictions that assumed a best-case scenario, 57% ($n=4$) still overpredicted (Supplementary File 1).

As of June 19th, 2020, there were over 50 studies that predicted the course of COVID-19. It is reassuring to have scientists rise to the challenge to help leaders make informed decisions. However, a review of forecasts from Ebola suggests that the majority of predictions were far from the eventual reality. In fact, COVID-19 predictions too have ranged from massive underestimates—a worst case scenario of

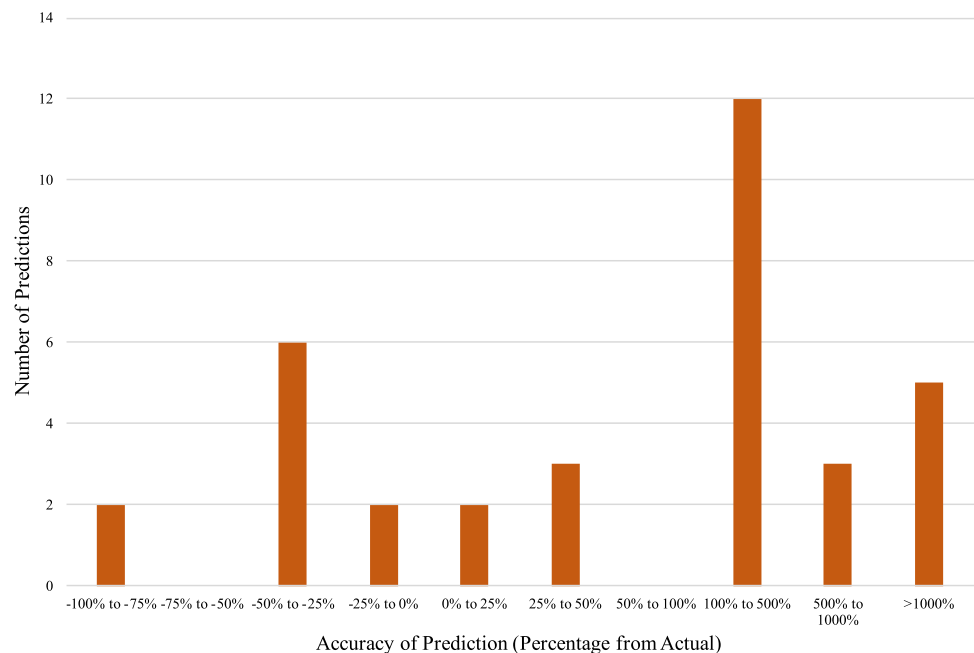
50,562 cases in Italy by May 31st (there were over 230,000 cases) [15, 16]—to proven overestimates—190,000 cases in Wuhan, China by April (there were 50,339 cases by the middle of May) [17, 18]. A model developed by the Institute for Health Metrics and Evaluation has gained prominence and has been widely utilized for state and federal policymaking. However, even this model has been often inaccurate at local and national levels, tending to provide overly narrow confidence intervals [19, 20]. Thus, we must consider the historically poor performance of disease prediction models when engaging with predictions for COVID-19.

Imperfect data, unverifiable assumptions, and the unpredictability of human behavior make forecasting epidemics an inherently uncertain task. For disease models to appropriately inform policy, we must acknowledge not only the uncertainty of prediction estimates (via confidence intervals), but also the uncertainty inherent to the exercise of prediction itself.

One approach for improving predictions is to incorporate a broader set of disciplinary perspectives. Often, disease forecasts are made on the basis of individual expertise in virology, infectious disease epidemiology, or demography. However, the psychology of how behavior changes, the economics of unemployment that ensues, and the policy options with which nations can respond also influence a pandemic's course but are typically left unconsidered. Models that integrate various forms of epidemic information will bring much needed nuance and humility to the challenge of prediction.

Furthermore, we recommend standardized reporting guidelines for forecasting studies, much like STROBE for observational epidemiological studies and CONSORT for randomized controlled trials. Forecasting studies should

Fig. 1 Frequency of predictions based on accuracy compared to actual numbers of deaths



discuss the “Current Forecasting Effort in Context” to summarize other predictions for the same outbreak as well as relevant predictions from prior outbreaks. These models should also report how their data was collected, detail the assumptions made and how realistic they are, and incorporate key epidemiological factors like age structure into the model [21]. Lastly, researchers should indicate how their forecast builds upon the existing landscape of predictions. As research and the media pivot focus towards the second surge of COVID-19, it is critical to quickly improve reporting standards so that future models are more honestly appraised.

Niels Bohr once said, “it is difficult to predict, especially the future”. Only once COVID-19 is behind us will we know whether prediction models did better than their counterparts from the Ebola epidemic. Until then, it is critical that researchers communicate the contexts and uncertainties of their predictions to best inform policy and the public.

References

- Powell A. Coronavirus screening may miss two-thirds of infected travelers entering U.S. *Harvard Gaz.* March 2020. <https://news.harvard.edu/gazette/story/2020/03/hundreds-of-u-s-coronavirus-cases-may-have-slipped-through-screenings/>. Accessed 7 July 2020.
- Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. London; 2020. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf>. Accessed 7 July 2020.
- Ganguly S. “India must prepare for a tsunami of coronavirus cases.” *BBC.* 2020. <https://www.bbc.com/news/av/world-asia-india-51962813/india-must-prepare-for-a-tsunami-of-coronavirus-cases>. Accessed 7 July 2020.
- Sturcke J. Bird flu pandemic “could kill 150M.” *The Guardian.* 2005. <https://www.theguardian.com/world/2005/sep/30/birdflu.jamessturcke#:~:text=%22The%20consequences%20in%20terms%20of,between%20five%20and%20150%20million%22>. Accessed 7 July 2020.
- Ferguson NM, Ghani AC, Donnelly CA, Hagens TJ, Anderson RM. Estimating the human health risk from possible BSE infection of the British sheep flock. *Nature.* 2002;415(6870):420–4. <https://doi.org/10.1038/nature709>.
- Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003–2015. World Health Organization. https://www.who.int/influenza/human_animal_interface/2020_MAY_tableH5N1.pdf?ua=1. Accessed 7 July 2020.
- Variant Creutzfeldt–Jakob disease Current Data (FEBRUARY 2015). The National CJD Research & Surveillance Unit. <https://web.archive.org/web/20150226031911/http://www.cjd.ed.ac.uk/documents/worldfigs.pdf>. Accessed 7 July 2020.
- Carias C, O’Hagan JJ, Gambhir M, Kahn EB, Swerdlow DL, Meltzer MI. Forecasting the 2014 West African Ebola Outbreak. *Epidemiol Rev.* 2019;41(1):34–50. <https://doi.org/10.1093/epirev/mxz013>.
- Ebola virus disease. World Health Organisation. <https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease>. Accessed 7 July 2020.
- Choi BCK, Pak AWP. A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths. *J Epidemiol Community Health.* 2003;57(10):831–35. <https://doi.org/10.1136/jech.57.10.831>.
- Cumulative Number of Reported Probable Cases Of SARS. World Health Organization. https://www.who.int/csr/sars/country/2003_05_31/en/. Accessed 7 July 2020.
- McNeil Jr. DG. Predicting Flu With the Aid of (George) Washington. *The New York Times.* 2009. <https://www.nytimes.com/2009/05/04/health/04model.html>. Accessed 7 July 2020.
- CDC Telebriefing on Investigation of Human Cases of H1N1 Flu May 15, 2009. <https://www.cdc.gov/media/transcripts/2009/t090515.htm>. Accessed 7 July 2020.
- Althaus CL. Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLoS Curr Outbreaks, Edition 1.* 2014. <https://doi.org/10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288>.
- Sotgiu G, Gerli GA, Centanni S, et al. Advanced forecasting of SARS-CoV-2 related deaths in Italy, Germany, Spain, and New York State. *Allergy.* 2020;75:1813–5. <https://doi.org/10.1111/all.14327>.
- Italy Records 75 New Coronavirus Deaths, 355 New Cases. *Reuters.* 2020. <https://www.reuters.com/article/us-health-coronavirus-s-italy-tally/italy-records-75-new-coronavirus-deaths-355-new-cases-idUSKBN2370JY>. Accessed 7 July 2020.
- Wan K, Chen J, Lu C, Dong L, Wu Z, Zhang L. When will the battle against novel coronavirus end in Wuhan: a SEIR modeling analysis. *J Glob Health.* 2020;10(1):011002. <https://doi.org/10.7189/jogh.10.011002>.
- Chen Y, Woo R. China’s Wuhan kicks off mass testing campaign for new coronavirus. *Reuters.* 13 May 2020. <https://www.reuters.com/article/us-health-coronavirus-china-wuhan/chinas-wuhan-kicks-off-mass-testing-campaign-for-new-coronavirus-idUSKBN22P1OD>. Accessed 7 July 2020.
- Miller K. Widely cited health institute keeps missing the mark on Maine death projections. *Portland Press Herald.* 2020. <https://www.pressherald.com/2020/06/18/widely-cited-model-keeps-missing-the-mark-on-maine-death-projections/>. Accessed 7 July 2020.
- Jewell NP, Lewnard JA, Jewell BL. Caution warranted: using the institute for health metrics and evaluation model for predicting the course of the COVID-19 pandemic. *Ann Intern Med.* 2020;14:M20–1565. <https://doi.org/10.7326/m20-1565>.
- Ioannidis JPA, Cripps S, Tanner M. Forecasting for COVID-19 has failed. *Int Inst Forecast.* 2020. <https://forecasters.org/blog/2020/06/14/forecasting-for-covid-19-has-failed/>. Accessed 7 July 2020.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.