



COMMENT



<https://doi.org/10.1057/s41599-020-00553-4>

OPEN

# Validity and usefulness of COVID-19 models

Sibel Eker  <sup>1</sup>✉

Mathematical models have become central to the public and policy debate about the recent COVID-19 pandemic. On the one hand, they provide guidance to policy-makers about the development of the epidemic and healthcare demand overtime; on the other hand, they are heavily criticized for their lack of credibility. This commentary reflects on three such models from a validity and usefulness perspective. Specifically, it discusses the complexity, validation, and communication of models informing the government decisions in the UK, US and Austria, and concludes that, although these models are useful in many ways, they currently lack a thorough validation and a clear communication of their uncertainties. Therefore, prediction claims of these models should be taken cautiously, and their merits on scenario analysis should be the basis for decision-making. The lessons that can be learned from the COVID models in terms of the communication of uncertainties and assumptions can guide the use of quantitative models in other policy-making areas.

<sup>1</sup>International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, 2361 Laxenburg, Austria. ✉email: [eker@iiasa.ac.at](mailto:eker@iiasa.ac.at)

## Introduction

The recent COVID-19 pandemic has drawn the attention of public, policy-makers, and scientific community one more time to the use of mathematical models in policy-making. Both public and policy-makers turned to modelers to answer questions like how many people would be infected and when the spread could end. In response to these questions, a plethora of hypothetical models are shared in online media platforms to inform the public debate about the underlying mechanisms of the outbreak, and a large number of scientific studies that project the epidemic dynamics in specific countries with better calibrated models have already been published or deposited in pre-publication repositories (Li et al., 2020; Kucharski et al., 2020; Wu et al., 2020; Massonnaud et al., 2020; Ghaffarzagdegan and Rahmandad, 2020). Open access data repositories have been rapidly established and have been enormously useful to track the situation and to inform the models (Xu et al., 2020a; Dong et al., 2020; Xu et al., 2020b). In addition, a few models are announced to assist the governments' decision-making in countries such as the UK, US, and Austria.

In the UK, the model developed by MRC Center for Global Infectious Disease Analysis at Imperial College London in collaboration with the World Health Organization (WHO) is often cited as the reason for the government to turn its policies to strict social distancing and lockdown (Boseley, 2020). The projections showed around 500,000 and 20,000 deaths in the UK without and with strict measures, respectively, and these different policy scenarios have been misinterpreted by the media as a drastic change in the model assumptions and raised questions about its accuracy (Sample, 2020).

In the US, the White House coronavirus task force announced in early April that they project between 100,000 and 240,000 deaths in the country based on an in-house model and the model developed by University of Washington's Institute for Health Metrics and Evaluation (IHME) (Bump, 2020). The accuracy of these models has been reported to be extensively discussed by policy-makers and scientists (Wan et al., 2020), and since projections change as new data is fed into the model, the debate about accuracy has been further sparked off (Tufekci, 2020; Jewell et al., 2020). Since then, the US Centers for Disease Control and Prevention (CDC) has compiled a large set of models from different institutions and this ensemble of projections by different models have been taken into account (CDC, 2020).

In Austria, the simulation model developed at Technical University of Vienna and partners (dwh, 2020) has been one of the main information sources for the government to formulate the outbreak strategy, while finding a wide coverage in the media (ORF, 2020; Aigner, 2020). As opposed to the UK and US, the accuracy of model projections did not create a heated debate in Austria, and the discussion focused on the relative impact of different scenarios rather than precise estimates of infection and casualty numbers.

Validity of the models used to assist decision-making has been discussed extensively for decades. A recent paper (Eker et al., 2018) has shown that the scientific literature on model validation in various fields relies heavily on data, and the practitioners' view supports this data-orientation since they consider the representativeness of a model and the accuracy of projections particularly important for validity. In other words, from a positivist point of view, validity equates with how good a model represents the reality, which is often measured by how accurately the model replicates the observed data. A commentary on this paper (Saltelli, 2019) highlighted the tradeoff between the representativeness of a model—the extent of complexity captured—and the propagation error caused by it, cautioning for overfocus on extending model boundaries hence creating a modeling “hubris”. Both

papers argue for a wider perspective on validity that goes beyond data-orientation and includes an explicit presentation of uncertainty and a participatory discussion on the usefulness of models, where usefulness refers to fit for a diverse set of purposes from assumption testing to systematic data compilation (Hodges, 1991).

In light of the above-mentioned two publications on validity, the purpose of this commentary is to reflect on three COVID-19 models that have been used in public policy. In particular, this paper briefly reviews how the three models capture the complexity of reality, how they report their validation, and how they communicate their assumptions and uncertainties. It then discusses how these models fit for their purposes and provides a future outlook on what the modeling community can learn from this experience.

## Models

The COVID-19 outbreak has created a typical post-normal science situation where facts are uncertain, stakes are high, and decisions are urgent (Funtowicz and Ravetz, 1995). The modeling studies are agile responses to this situation synthesizing the best available knowledge in a limited time, and they have not been subject to peer-review yet. This commentary acknowledges the situation and takes the non-peer reviewed documentation of the three studies into account. Table 1 summarizes the main features of these models, their reported validation and uncertainty communication.

**UK.** The COVID-19 model of the MRC Center for Global Infectious Disease Analysis at Imperial College London is a mechanistic hierarchical Bayesian model that links policy interventions, infection and death rates with Bayesian probabilities fitted into the observed number of deaths. It is conceptually based on the Susceptible-Infected-Recovered (SIR) framework of epidemiological modeling, yet it is not expressed in differential form to account for the rates of change and accumulation explicitly. As stated in the report from mid-March (Flaxman et al., 2020), which the discussion in this commentary is based on, the purpose of the model is “to infer the impact of these [policy] interventions across 11 European countries”, hence an ex-post policy analysis. A secondary purpose is to back-calculate the number of actual infections, which are expected to be much higher than the detected number of cases. The model focuses on the epidemical process of transmission, infection and casualties, therefore the nonlinearities caused by social behavior, healthcare system capacity and treatment choices are not explicitly taken into account. This choice of scope can be argued to align with the stated purpose of the model and to help avoiding the model propagation error.

The reporting clearly highlights the key assumptions of the model and tests the implications of some of these assumptions with sensitivity analyses. For instance, since the model is continuously updated according to the available data, it can be biased towards the countries hit by the outbreak earlier and produced more data, such as China, Italy, and Spain. This assumption is tested with a leave-out analysis, where the data of Italy and Spain, one at a time, is excluded from the dataset used for fitting the model. The results did not differ significantly, therefore led to the conclusion that there is no strong dependence on any one country in the model. Another key assumption is fitting the model only to the death data, although this implies that many intermediary parameters that do not directly relate to the death rate are estimated according to it. The authors justify this choice by arguing that the death data is the most reliable one to

**Table 1 Summary of the three Covid-19 models assisting public policy.**

	<b>Complexity (representation)</b>	<b>Validation</b>	<b>Communication</b>
UK	<ul style="list-style-type: none"> <li>• Mechanistic Bayesian hierarchical model, where transmission rate is fitted into the death data</li> <li>• Model scope limited to the transmission dynamics, i.e., infections, recoveries and deaths, without an explicit consideration of the healthcare system</li> <li>• Interventions are the only factors changing the basic reproductive number (<math>R</math>), and the intervention effect on each country is assumed to be the same</li> </ul>	<ul style="list-style-type: none"> <li>• The stated purpose is ex-post analysis of the impact of interventions, and to calculate the number of actual infections</li> <li>• Cross-validation of model-generated death rate with the observed data. Strong conclusions based on limited testing</li> <li>• Convergence of probability distribution estimates verifying that the model is computationally sound within its framework</li> <li>• Sensitivity analysis to various assumptions, yet not discussed in terms of the eventual deaths</li> </ul>	<ul style="list-style-type: none"> <li>• Replication code is publicly available</li> <li>• Clear statement of assumptions throughout the report, yet limited discussion about the implications, i.e., how the results could be different, except the sensitivity analyses</li> <li>• Predictive tone emphasizing the forecasting purpose of the model</li> </ul>
US	<ul style="list-style-type: none"> <li>• Nonlinear mixed effects model fitting the cumulative death rate to a sigmoidal function (Gaussian error function)</li> <li>• Back-computing the hospital bed and ICU demand for given death rates for different age cohorts</li> <li>• Useful estimates for excess hospital demand for each state in the US</li> </ul>	<ul style="list-style-type: none"> <li>• The stated purpose is “to determine the extent and timing of deaths and excess demand for hospital services due to Covid-19 in the US”</li> <li>• No explicitly reported validation tests. Data use, e.g., sources and description of the model remains as the only info about the credibility of the model</li> </ul>	<ul style="list-style-type: none"> <li>• Model results are publicly shared with an informative visualization tool (IHME, 2020), and key assumptions are summarized in FAQ</li> <li>• Long-term (4 months) forecasts are claimed for the peak hospital demand and peaking time, without an explicit validation</li> </ul>
Austria	<ul style="list-style-type: none"> <li>• Extension on an existing agent-based population model</li> <li>• Highly detailed, individual level model with explicit representation of various types of social contacts, hence allowing for detailed comparison of lockdown and social distancing policies</li> <li>• Due to the level of detail, potentially suffering from propagation error</li> </ul>	<ul style="list-style-type: none"> <li>• No reported validation tests for the Covid-19 extension although the underlying population model is validated</li> <li>• The only information on credibility is a conceptual description of the model and the data sources</li> </ul>	<ul style="list-style-type: none"> <li>• Nether the model code nor the projections are publicly available</li> <li>• Admittance of uncertainties and disclaimer that the model assists in assessing semi-quantitative relative impact of interventions</li> </ul>

reflect the actual situation, since the actual number of infections is highly uncertain beyond the reported cases. They show the decreasing trend of log-linear death rates (decreasingly increasing number of deaths) to illustrate that the model estimates about the post-intervention reproductive number  $R$  (the number of cases caused by each case in a susceptible population after lockdown measures) are driven by the data, not by strong model assumptions. This can be considered as an evidence of capturing the correct impact of interventions conceptually, yet not numerically.

The reported validation of the model includes the comparison of the 3-day model forecasts to the observed data with a good match between the two. The comparison is on a logarithmic scale, therefore misleading for the match of actual predicted and observed death rates. The reader is also not informed about whether this comparison is made on a country basis or over time. In terms of communication of this validation test, the outcome is presented with a high certitude, such as “a strong empirical justification” for the plausibility and appropriateness of the model estimates. A second validation test is conducted to check the convergence of Markov Chain Monte Carlo simulations used for parameter fitting. The resulting good convergence verifies that the model computations are reliable within its framework, yet it does not directly contribute to validating the model with respect to the real life phenomena and its fit for purpose.

**US.** The model developed at University of Washington’s Institute for Health Metrics and Evaluation (IHME) is also a statistical model, specifically a nonlinear mixed effects model fitted to the available data for cumulative deaths in China, Italy, and the US (Murray, 2020). Its purpose is specified as determining the extent

and timing of deaths and excess demand for hospital services in the US, later extended to different countries. Therefore, the model does not capture the epidemic’s transmission dynamics, and focuses only on forecasting the death rate and the hospitalization demand inferred from it. This relatively narrow scope can imply an inadequacy error described by Saltelli (2019) for reliable projections for the peak and duration of the epidemic, yet by addressing a pragmatic problem, i.e., the healthcare system capacity, it provides useful estimates for decisionmakers at the federal or state level about what can be done to minimize the fatalities. The IHME model projections differ from other models (CDC, 2020), especially from the UK model by MRC Center for Global Infectious Disease Analysis. For instance, the former predicts the number of daily deaths to decrease from 108 to 90 in the first week of July (IHME, 2020) in the UK, whereas the latter estimates it to be relatively stable around 200 (MRC-IDE, 2020).

The documentation of the study when this commentary was prepared at the end of March did not report any explicit validation or sensitivity tests. A comparison of model forecasts and the observed data is missing. The only reported information that can be used to assess the model reliability is the key assumptions and data sources fed into the model. As for the projections of the model, a clear and concise communication strategy is adopted with an online dashboard that publicly shares the estimated death rates and hospital capacity needs over time.

**Austria.** Unlike the previous two prognostic statistical models, the Austrian COVID-19 model is a descriptive simulation model that is extended from an existing population model (Bicher et al., 2018) based on the SIR framework. It follows a highly detailed agent-based modeling approach, where each individual in the

society and various contact places, such as schools, kindergartens, leisure places, are explicitly modeled (dwh, 2020). The model does not back-calculate the post-intervention reproductive number of the epidemic ( $R$ ), but generates it by simulating the social contact reduction. Therefore, it allows for a detailed ex-ante analysis of various social distancing and lock-down interventions. However, the projections are not shared in detail in the current publicly available documentation. Only semi-quantitative description of the results is available (Emrich et al., 2020).

Regarding the validation, the underlying population model is stated to be thoroughly validated, yet there is no reporting to date about the validation of the COVID-19 extension. Besides the usefulness to assess the impact of social distancing policies, it can be expected to suffer from model propagation error (Saltelli, 2019) due to the assumptions required to calibrate such a large and detailed model. Although no quantified measure of this error is shared in the available discussion of the projections, the team acknowledges the uncertainties with a disclaimer that “the results must not be read as absolute figures” and emphasize the role of models as heuristic tools that can help “asses the effects and orders of magnitude” (dexhelpp, 2020) in different scenarios.

## Discussion

The three models discussed in this paper are undeniably useful in several aspects. They inform the public and policy debate, alongside other tools, about the extent of the epidemic, the scale of the healthcare problem, and the relative effectiveness of intervention strategies. They synthesize and process the best available knowledge and data, and provide a structured framework to understand and interpret the epidemical and social mechanisms behind the outbreak. They fit well for the purpose of testing our assumptions about the uncertainties surrounding the outbreak and creating a variety of what-if scenarios.

However, they cannot be considered as accurate prediction tools, not only because no model is able to this, but also because these models are lacking thorough formal validation according to their current reports. Statistical models, which are inherently short-term forecasting tools, are useful in the above-mentioned aspects, yet they cannot be used for months-ahead forecasts while their ability is tested only for a few days. Besides the misinterpretation by the media that triggered the public debate about accuracy, expressions of overconfidence are also present in the official reports of the models even though quantitative measures of uncertainty and prediction error are scarcely mentioned. Modeling practitioners, especially the experienced ones, acknowledge that decisionmakers find a model credible if the assumptions and uncertainties are communicated well (Eker et al., 2018). Such clear communication can increase the credibility of COVID-19 models, too.

The pandemic created a situation where facts are uncertain, stakes are high, and decisions are very urgent, which is similar to many policy-making cases from climate change mitigation to sustainable resource management. Therefore, the lessons learned from the case of COVID models, for instance in terms of data use, communication of uncertainties and assumptions, and informing policy-makers, can resonate in other disciplines that provide policy recommendations based on quantitative models. During the pandemic, the urgency for decision-making did not leave room for extensive peer-review and auditing of the models. Still, post-crisis research can analyze the usefulness of these models in the discourse and decision-making, for instance by surveying the decisionmakers and modelers about the challenges of model use in this particular science-policy interface. In that way, not only the epidemiological models can be better prepared for the next

outbreak, but also the policy modeling community can learn from this high-stake and high-uncertainty situation. Until then, prediction claims of any model should be taken with a pinch of salt, and their merits on scenario analysis should guide decision-making.

Received: 14 April 2020; Accepted: 15 July 2020;

Published online: 03 August 2020

## References

- Aigner F (2020) COVID-19: model calculations by TU Wien Informatics confirmed. <https://informatics.tuwien.ac.at/news/1873>
- Bicher M, Urach C, Popper N (2018) Gepoc ABM: a generic agent-based population model for Austria. 2018 Winter Simulation Conference (WSC). IEEE, pp. 2656–2667
- Boseley S (2020) New data, new policy: why UK’s coronavirus strategy changed. *The Guardian*
- Bump P (2020) The grim death-toll projections the White House offered Monday have already been revised upward. *The Washington Post*
- CDC (2020) Forecasts of total deaths. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>
- dexhelpp (2020) COVID-19: measures to show effect soon. <http://dexhelpp.at/en/news/covid-19-measures-to-show-effect-soon/>
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20:533–534
- dwh (2020) Simulation of the SARS-CoV-2 epidemic in Vienna. dwh simulation Services, Vienna
- Eker S, Rovenskaya E, Obersteiner M et al. (2018) Practice and perspectives in the validation of resource management models. *Nat Commun* 9:5359
- Emrich Š, Bicher M, Urach C et al. (2020) Corona: worst case scenario and counter-measures. <http://dexhelpp.at/en/news/corona-worst-case-scenario-and-counter-measures/>
- Flaxman S, Mishra S, Gandy A et al. (2020) Report 13: estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College COVID-19 Response Team
- Funtowicz SO, Ravetz JR (1995) Science for the post normal age. In: Westra L, Lemons J (eds) *Perspectives on ecological integrity*. Springer, pp. 146–161
- Ghaffarzadegan N, Rahmandad H (2020) Simulation-based estimation of the spread of COVID-19 in Iran. <https://doi.org/10.1101/2020.03.22.20040956>
- Hodges JS (1991) Six (or so) things you can do with a bad model. *Oper Res* 39:355–365
- IHME (2020) COVID-19 projections. University of Washington
- Jewell NP, Lewnard JA, Jewell BL (2020) Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA* 323:1893–1894
- Kucharski AJ, Russell TW, Diamond C et al. (2020) Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 20:553–558
- Li R, Pei S, Chen B et al. (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* 368:489–493
- Massonnaud C, Roux J, Crépey P (2020) COVID-19: forecasting short term hospital needs in France. <https://doi.org/10.1101/2020.03.16.20036939>
- MRC-IDE (2020) Short-term forecasts of COVID-19 deaths in multiple countries. <https://mrc-ide.github.io/covid19-short-term-forecasts/index.html#authors>
- Murray CJ (2020) Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. <https://doi.org/10.1101/2020.03.27.20043752>
- ORF (2020) TU-Experten simulieren mehrere Scenarien. <https://science.orf.at/stories/3200425/>
- Saltelli A (2019) A short comment on statistical versus mathematical modelling. *Nat Commun* 10:3870
- Sample I (2020) Coronavirus exposes the problems and pitfalls of modelling. *The Guardian*
- Tufekci Z (2020) Don’t Believe the COVID-19 Models—that’s not what they’re for. <https://www.theatlantic.com/technology/archive/2020/04/coronavirus-models-arent-supposed-be-right/609271/>
- Wan W, Dawsey J, Parker A et al. (2020) Experts and Trump’s advisers doubt White House’s 240,000 coronavirus deaths estimate. *The Washington Post*
- Wu JT, Leung K, Leung GM (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 395:689–697
- Xu B, Gutierrez B, Mekaru S et al. (2020a) Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* 7:1–6

Xu B, Kraemer MU, Gutierrez B et al. (2020b) Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis* 20:534

### Competing interests

The author declares no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.E.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020