

Preprint

Transparency Assessment of COVID-19 Models

Summary: *Evaluation of 29 impactful COVID-19 models reveals inconsistent adherence to best transparency practices; higher transparency is needed to inform policy.*

Mohammad S. Jalali^{1,2*}, Catherine DiGennaro¹, Devi Sridhar^{3,4}

¹MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, USA

²MIT Sloan School of Management, Cambridge, MA, USA

³Medical School, The University of Edinburgh, Edinburgh, UK

⁴Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, UK

*Correspondence to: msjalali@mgh.harvard.edu

Abstract: As the COVID-19 pandemic has caused major societal unrest, modelers have worked to project future trends of COVID-19 and predict upcoming challenges and impacts of policy action. These models, alone or in aggregate, are influential for decision-makers at every level. Therefore, the method and documentation of COVID-19 models must be highly transparent to ensure that projections and consequential policies put forth have sound epistemological grounds. We evaluated 29 COVID-19 models receiving high attention levels within the scientific community and/or informing government responses. We evaluated these models against 27 transparency criteria. We found high levels of transparency in model documentation aspects such as reporting uncertainty analysis; however, about half of the models do not share code and a quarter do not report equations. These discrepancies underscore the need for transparency and reproducibility to be at the forefront of researchers' priorities, especially during a global health crisis when stakes are critically high.

Main text: The COVID-19 pandemic has strained societal structures and created a global crisis. Scientific models play a critical role in mitigating the pandemic's harm, from estimating the spread of outbreaks to analyzing the effects of public health policies. Given the gravity of this crisis, the context- and time-sensitive measures with real population health impacts that COVID-19 models provide are of utmost importance. These models must be completely transparent before policies and insights are enacted.

Transparency is the cornerstone of the scientific method and efforts to improve transparency and reproducibility of research have been increasing over the past few years (1). Recently, *Science* called for complete transparency of COVID-19 models (2). Lack of such transparency in the design, development, and analysis of these models not only reduces the trust in their timely messages, but also limits the reproducibility of the models, impeding other scientists from verifying the findings and improving a model's performance via further explorations and innovation. Many modelers have already shared the details of their models openly, yet the overall status of transparency of COVID-19 models remains unknown.

To systematically evaluate the transparency of COVID-19 models, we reviewed a sample of models that have earned global attention and been widely used to inform public health policies. We first collected COVID-19 models that included a methods write-up from CDC's compilation (3), then collected the most-cited COVID-19 models in Google Scholar; this resulted in 29 models for evaluation. Due to the urgency of the pandemic, preprints and project websites made available in advance of publication and have had an essential role during the crisis (4), and therefore, we included models from these sources (n=12) in addition to peer-reviewed publications (n=17).

We assessed these sample models against 27 binary criteria to evaluate the transparency of their reports. Adopted from several transparency checklists (5-7) and tailored to evaluate models, two main themes guide the transparency assessment criteria: 1) specificity of model items, including but not limited to discussion of model mechanisms, assumptions, parametrization, formulation, codes, and sensitivity analysis; and 2) general research items, such as disclosure of research limitations, funding, and potential conflict of interest. Two trained researchers reviewed the full text and appendix of each modeling report and a third reviewer helped to discuss the discrepancies.

The results of our evaluation are reported in Figure 1. On average, the transparency criteria are satisfied by 75% of the sample models. While eight of the criteria are satisfied by 90% of the models, most of the criteria are satisfied by a much smaller percent of models. For instance, 21% of the models do not report the sources of their longitudinal data, 24% do not report their equations, 31% do not report their estimated parameters, 48% do not share their longitudinal data, and 52% do not report their code. Among the sample articles, only four articles satisfied 90% of our transparency checklist items.

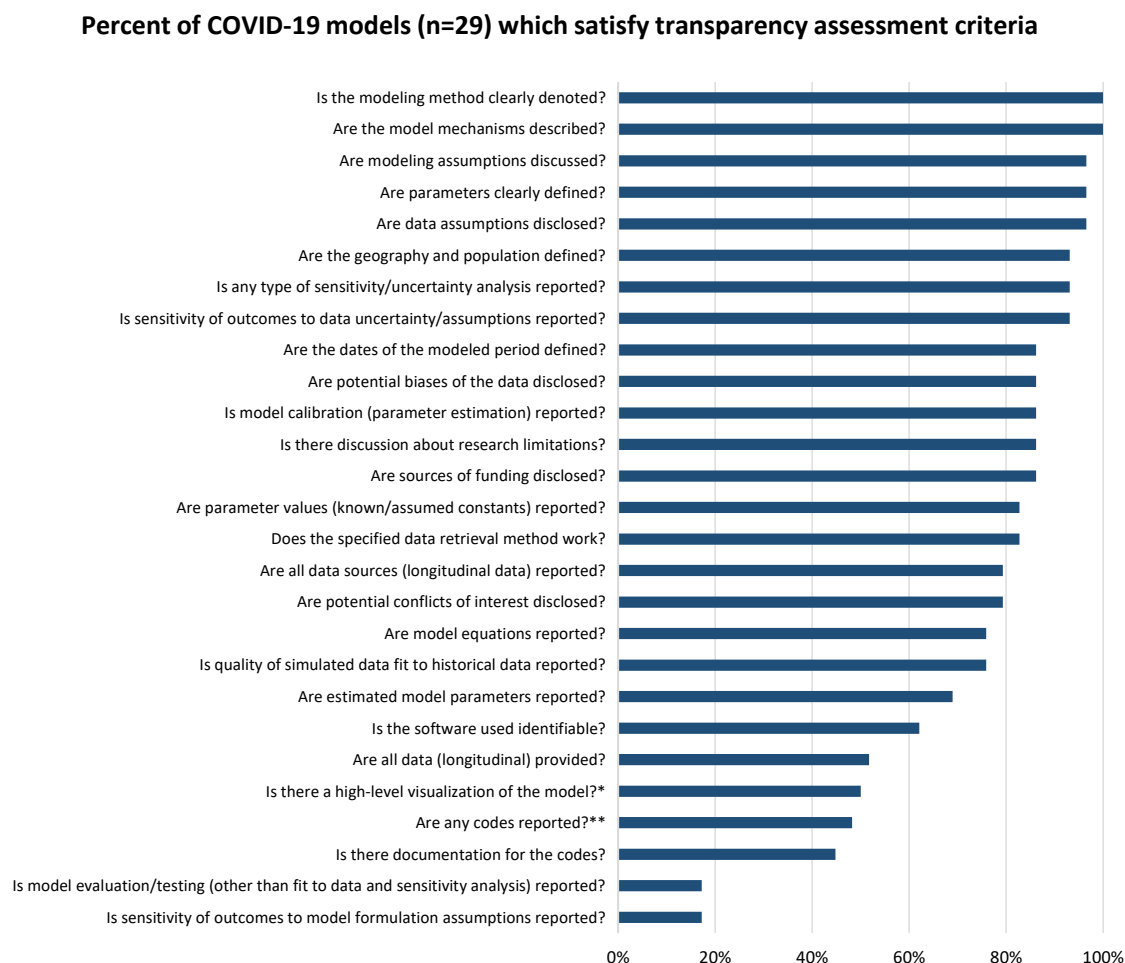


Fig. 1. Percent of COVID-19 models (n=29) which satisfy transparency assessment criteria

*Five regression models are exempt from this criterion, as there is no visualization that communicates model structure. **codes used for a generalized model are not sufficient; we were able to successfully retrieve the codes of each model that provided a retrieval method.

Evaluations like this one demonstrate that a model which is not fully transparent can still posit analytical insights and propose policy. Rather than presenting recommendations at face value, it

is imperative for modelers to make sure that their claims are able to be independently verifiable. The scientific and modeling communities can and must hold themselves accountable to make transparency the norm, not the exception, or else risk losing the faith of policymakers and the public.

Such consequences were observed when IHME released their model in late March with highly criticized projections; they presented confidence intervals that converged in the future and seemingly low projections of case numbers and deaths during the first wave, among other things, and the scientific community was further frustrated by the lack of codes and other model details. This model was often cited by government agencies, including the White House (8), and its negative reception was difficult to overcome, even after their release of ostensibly more realistic projections. Another high-profile preprint model, published March 16, 2020 by Imperial College London, experienced similar scrutiny after predicting 510,000 COVID-19 deaths in a no-intervention scenario, prompting researchers to attempt replication. Unfortunately, the research team withheld their code for nearly six weeks following publication. Meanwhile, the United Kingdom instituted the recommended stringent stay-at-home measures. When the code was released in late April, several bugs and assumptions were unearthed and was not until June that the replication attempts were successful (9).

A crucial element of transparency is achieved by providing codes. One concern for researchers is that their code is disorganized and cannot be evaluated or run by laypeople hoping to replicate their efforts. However, this is a misconception; even messy code can provide a framework for an accurate replication and generate useful dialogue, as seen on platforms like GitHub. Still, well-documented code is preferable. Of the 48% of articles which reported their codes, all but one provided helpful detailed documentation either directly in the file or in a supplement. We encourage COVID-19 modelers who hope to impact perceptions and policy to release their codes in a timely manner for public evaluation.

Many journals ask for transparency statements and encourage scientists to report the details in supplemental documents. Research shows the data sharing policies of journals has increased the frequency and quality of data sharing altogether (10). While journals need to further enhance their publication policies and increase their transparency requirements, journals' options are limited, and they cannot control the full transparency of publications. During a crisis such as

COVID-19, preprints provide speedy information delivery before peer review, therefore, journal requirements and policies make minimal impact. Models which were still preprints or project websites satisfied an average of 70% of the transparency criteria, as compared to peer-reviewed articles' 79%. The responsibility of transparency remains largely on the shoulders of modelers, even though the peer-review process can help address these omissions. It is imperative that modelers take it upon themselves to follow open research practices, adopt documentation and reporting guidelines, and share full details of their models.

Reporting a fully documented and transparent model can be difficult, but this effort has both tangible and intangible benefits for the modelers. With the urgency of a global pandemic, modelers might justify putting transparency second to the speed of reporting, however, poor transparency of models that directly impact public health policies and therefore human lives can have major harmful consequences. Hence, *all* models must be fully transparent for both scientific and ethical purposes.

References and Notes:

1. M. McNutt, Journals unite for reproducibility. *Science* **346**, 679-679 (2014).
2. C. M. Barton *et al.*, Call for transparency of COVID-19 models. *Science* **368**, 482-483 (2020).
3. Centers for Disease Control and Prevention, "Forecasts of Total Deaths," (2020).
4. M. S. Majumder, K. D. Mandl, Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. *The Lancet Global Health* **8**, e627-e630 (2020).
5. T. E. Hardwicke *et al.*, An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014-2017). *Royal Society Open Science* **7**, 190806 (2020).
6. G. A. Stevens *et al.*, Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *The Lancet* **388**, e19-e23 (2016).
7. J. D. Wallach, K. W. Boyack, J. P. A. Ioannidis, Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLOS Biology* **16**, e2006930 (2018).
8. M. Shear, M. Crowley, J. Glanz, Coronavirus may kill 100,000 to 240,000 in US despite actions, officials say. *New York Times* **31**, (2020).
9. D. S. Chawla, Critiqued Coronavirus Simulation Gets Thumbs Up From Code-Checking Efforts. *Nature* **582**, 323-324 (2020).
10. T. E. Hardwicke *et al.*, Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science* **5**, 180448 (2018).

Funding: No funding was used to conduct this study.

Competing interests: Authors declare no competing interests.

Data and materials availability: All analysis details are available in the supplementary materials.

Table S1: Transparency Assessment of 29 COVID-19 models

Transparency Assessment Criteria	COVID-19 Model																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Are the geography and population defined?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Are the dates of the modeled period defined?	Y	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Is the modeling method clearly denoted?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Are the model mechanisms described?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Is there a high-level visualization of the model?*	N	N	Y	Y	N	E	Y	E	Y	Y	N	E	N	N	Y	Y	Y	Y	E	Y	Y	N	N	N	N	N	N	Y	E
Are model equations reported?	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Are modeling assumptions discussed?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Are parameter values (known/assumed constants) reported?	Y	Y	Y	Y	Y	Y	Y	P	P	Y	Y	N	P	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Are all data sources (longitudinal data) reported?	P	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	P	Y	Y	Y	Y	Y	Y	P	P	Y	Y	Y	Y
Are all data (longitudinal) provided?	P	Y	Y	Y	N	P	Y	N	Y	N	Y	P	Y	Y	Y	N	Y	Y	N	N	P	Y	N	P	N	Y	Y	Y	N
Does the specified data retrieval method work?	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	P	Y	Y	Y	Y	P	Y	Y	Y
Are parameters clearly defined?	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Are data assumptions disclosed?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Are potential biases of the data disclosed?	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N
Are any codes reported?*	N	N	Y	Y	N	N	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	N	N	N	Y	N
Is there documentation for the codes?	N	N	Y	Y	N	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	N	Y	N	N	N	Y	N	N	N	Y	N
Is the software used identifiable?	N	N	Y	Y	N	N	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	N
Is model calibration (parameter estimation) reported?	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y
Are estimated model parameters reported?	N	Y	Y	Y	N	N	Y	N	Y	Y	Y	N	N	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y
Is quality of simulated data fit to historical data reported?	N	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y
Is model evaluation/testing (other than fit to data and sensitivity analysis) reported?	N	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N	Y	N	Y	N	N	N	N	N	N
Is any type of sensitivity/uncertainty analysis reported?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Is sensitivity of outcomes to data uncertainty/assumptions reported?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Is sensitivity of outcomes to model formulation assumptions reported?	N	N	Y	N	N	Y	N	N	N	N	Y	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	Y	N	N
Is there discussion about research limitations?	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y
Are sources of funding disclosed?	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
Are potential conflicts of interest disclosed?	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

*n=24, there was no expectation of this for purely statistical models; **we were also able to retrieve all codes using the specified method.

Y=Yes; N=No; P=Partially; E=Exempt only applicable for the assessment of high-level model visualization

Reviewed models:

1. Wu, J.T., K. Leung, and G.M. Leung, *Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study*. Lancet, 2020. **395**(10225): p. 689-697.
2. Branas, C.C., et al., *Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus SARS-CoV2) cases in US counties*. medRxiv, 2020.
3. Flaxman, S., et al., *Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe*. Nature, 2020.
4. Walker, P., et al., *Report 12: The global impact of COVID-19 and strategies for mitigation and suppression*. 2020.
5. Ferguson, N., et al., *Report 9: Impact of non-pharmaceutical interventions NPIs) to reduce COVID19 mortality and healthcare demand*. 2020.
6. Murray, C., *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*. 2020.
7. Chinazzi, M., et al., *The effect of travel restrictions on the spread of the 2019 novel coronavirus COVID-19) outbreak*. Science, 2020. **368**(6489): p. 395-400.
8. Osthus, D.D.V., S. *COVID-19 Confirmed and Forecasted Case Data*. 2020; Available from: <https://covid-19.bsvgateway.org/>.
9. Bertsimas, D., *DELPHI Epidemiological Model Documentation*. 2020.
10. Moss, R., et al., *Modelling the impact of COVID-19 in Australia to inform transmission reducing measures and health system preparedness*. medRxiv, 2020.
11. Li, R., et al., *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus SARS-CoV-2)*. Science, 2020. **368**(6490): p. 489-493.
12. Spencer Woody, M.T., et al., *Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones*. 2020.
13. Gu, Y. *COVID-19 Projections Using Machine Learning*. 2020; Available from: <https://covid19-projections.com/>.
14. España, G., *Forecasting COVID-19 mortality in the US midwest*. 2020.
15. Kucharski, A.J., et al., *Early dynamics of transmission and control of COVID-19: a mathematical modelling study*. The lancet infectious diseases, 2020.
16. Hellewell, J., et al., *Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts*. The Lancet Global Health, 2020.
17. Prem, K., et al., *The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study*. The Lancet Public Health, 2020.
18. Kissler, S.M., et al., *Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period*. Science, 2020. **368**(6493): p. 860-868.
19. Roosa, K., et al., *Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020*. Infectious Disease Modelling, 2020. **5**: p. 256-263.
20. Peng, L., et al., *Epidemic analysis of COVID-19 in China by dynamical modeling*. arXiv preprint arXiv:2002.06563, 2020.
21. Yang, Z., et al., *Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions*. Journal of Thoracic Disease, 2020. **123**): p. 165.
22. Lin, Q., et al., *A conceptual model for the outbreak of Coronavirus disease 2019 COVID-19) in Wuhan, China with individual reaction and governmental action*. International journal of infectious diseases, 2020.
23. Fanelli, D. and F. Piazza, *Analysis and forecast of COVID-19 spreading in China, Italy and France*. Chaos, Solitons & Fractals, 2020. **134**: p. 109761.
24. Boldog, P., et al., *Risk assessment of novel coronavirus COVID-19 outbreaks outside China*. Journal of clinical medicine, 2020. **92**): p. 571.
25. Chang, S.L., et al., *Modelling transmission and control of the COVID-19 pandemic in Australia*. arXiv preprint arXiv:2003.10218, 2020.
26. Shen, M., et al., *Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China*. BioRxiv, 2020.
27. Zhao, S. and H. Chen, *Modeling the epidemic dynamics and control of COVID-19 outbreak in China*. Quantitative Biology, 2020. **81**): p. 11-19.
28. Davies, N.G., et al., *The effect of non-pharmaceutical interventions on COVID-19 cases, deaths and demand for hospital services in the UK: a modelling study*. medRxiv, 2020: p. 2020.04.01.20049908.
29. Peak, C.M., et al., *Comparative Impact of Individual Quarantine vs. Active Monitoring of Contacts for the Mitigation of COVID-19: a modelling study*. medRxiv, 2020: p. 2020.03.05.20031088.