Purposes and duties in scientific modelling

Eric Winsberg ,[1] Stephanie Harvard [2]

[1]Department of Philosophy, University of South Florida, Tampa, Florida, USA
[2]Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, British Columbia, Canada

Correspondence to
Dr Eric Winsberg, University of South Florida, Tampa, Florida, USA; winsberg@usf.edu

ABSTRACT

More people than ever are paying attention to philosophical questions about epidemiological models, including their susceptibility to the influence of social and ethical values, sufficiency to inform policy decisions under certain conditions, and even their fundamental nature. One important question pertains to the purposes of epidemiological models, for example, are COVID-19 models for 'prediction' or 'projection'? Are they adequate for making causal inferences? Is one of their goals, or virtues, to change individual responses to the pandemic? In this essay, we offer our perspective on these questions and place them in the context of other recent philosophical arguments about epidemiological models. We argue that clarifying the intended purpose of a model, and assessing its adequacy for that purpose, are moral-epistemic duties, responsibilities which pertain to knowledge but have moral significance nonetheless. This moral significance, we argue, stems from the inherent value-ladenness of models, along with the potential for models to be used in political decision making in ways that conflict with liberal values and which could lead to downstream harms. Increasing conversation about the moral significance of modelling, we argue, could help us to resist further eroding our standards of democratic scrutiny in the COVID-19 era.

INTRODUCTION

More people than ever are paying attention to epidemiological models, including philosophers of science. A result is a greater focus on philosophical questions about this type of modelling, such as its susceptibility to the influence of social and ethical values,[1] its sufficiency to inform policy decisions under certain conditions,[2] and even its fundamental nature—as Fuller[3] recently asked 'what are COVID-19 epidemic models modelling, anyway?'. Although Fuller's question is multifaceted, an important part of it pertains to model purposes: indeed, there is considerable confusion about how exactly certain COVID-19 models are meant to be used. Are these models for 'prediction' or 'projection'?[3 4] Are they adequate for making causal inferences?[3] Is one of their goals, or virtues, to change individual responses to the pandemic?[5] In this essay, we offer our perspective on these questions, first by giving our own answers to them, then by placing them in the context of other recent philosophical arguments about epidemiological models.[1 3 6] Specifically, we argue that (1) clarifying the intended purpose of a model and (2) assessing its adequacy for that purpose are ongoing moral-epistemic duties that must be upheld throughout the modelling process. In other words, certain modelling tasks are moral responsibilities, duties which pertain to knowledge but are moral nonetheless. The moral significance of modelling, we argue, stems from the inherent value-ladenness of models,[1 7 8] along with the potential for models to be used in political decision making in ways that conflict with liberal values[2 9] and which could lead to downstream harms. Increasing conversation about the moral significance of modelling, we argue, could help us to resist further eroding our standards of democratic scrutiny in the COVID-19 era.[10]

PREDICTION, PROJECTION OR POSSIBLY MORE?

Two models that have gained much attention during the COVID-19 pandemic are the Imperial College London (ICL) model[11] and Institute for Health Metrics and Evaluation (IMHE) model.[12] Briefly, the main goal of the ICL model was to represent the impact of various non-pharmaceutical interventions on COVID-19 deaths and hospital/intensive care unit admissions in Great Britain and the USA (although the authors note that results are 'equally applicable to most high-income countries'(Ferguson *et al*, p4).[11] The IMHE model aimed to describe the need for hospital beds, intensive care unit beds, and ventilators due to COVID-19 based on projected deaths for all US states.[12] From a technical perspective, the two models are very different: the ICL is an individual-based simulation model, while the IMHE projects deaths using a statistical model. For the purpose of our discussion, these technical differences are of limited importance; we will note them as we go along.

Prediction versus projection

Recently, Fuller[3] and Schroeder[4] have analysed the ICL and/or IHME models with the same question in mind: what kinds of predictions do these models make? Following Fuller[3] and Schroeder,[4] there are two possibilities to consider. The first is that these models make predictions about what will actually occur, the second that they make predictions about what would happen under certain conditions. Both Fuller[3] and Schroeder[4] call the first type of prediction forecasting and the second conditional projection. As each writer gives just a brief rundown of the distinction, we will give our own, more detailed explanation of it following Winsberg.[8] Winsberg[8] describes how the distinction between forecasting (or just 'prediction') and projection is used in climate science, but, as we will show, his discussion can be fruitfully applied to epidemiology. Distinguishing between prediction and projection helps us to see that different types of models invite us to use different strategies to assess their adequacy for purpose, which is relevant to debates

surrounding different conceptions of 'model validation' in the health sciences.[13]

As Winsberg[8] explains, in climate science, the difference must often be stressed between weather models and climate models. Weather models are indeed used to tell us what will actually occur at a particular time and place, that is, to make spatially and temporally fine-grained predictions about states of the atmosphere. Weather models have three important features: (1) they take into account our best and most recent measurements of the current atmospheric conditions in the relevant region; (2) their predictions are based on a comparatively high degree of fidelity to our near-perfect understanding of causal relationships in weather systems; (3) the accuracy of their predictions is easy to test empirically: they make frequent predictions concerning the near future that we soon come to observe in great detail. When a model has these features like these, our strategy for assessing its adequacy for purpose can be relatively straightforward, focused on the match between our predictions and our later observations.

Climate models, on the other hand, do not make spatially and temporally fine-grained predictions about states of the atmosphere. The chaotic nature of the atmosphere makes this impossible to do on the century-long time scales that concern climate science. Rather, climate models project global or wide-regional averages of the variables describing these states, averaged over timespans of 30 years or longer. Furthermore, rather than forecast what will actually happen in the future, climate models are generally used to explore what possible evolutions in the climate could be triggered by different external forcings, defined as perturbations that are outside the climate system but capable of pushing it beyond its normal range of variation (eg, ozone depletion, C02 emissions, other greenhouse gases, deforestation). Climate models themselves have three important features: (1) they model climate variables in the form of coarse-grained statistical summaries of weather variables (eg, averages, degrees of variation in weather variables) and do not take account of current atmospheric conditions; (2) their projections are based on a representation of the dynamics of the atmosphere and ocean that, in order to be computationally manageable, has a much lower degree of fidelity to our best physics than weather prediction models; (3) the accuracy of their predictions is much less testable than those of weather models. This reduced testability is because we only get one run of the planet's climate evolution and we will not see it until it is too late: although climate models are meant to project the effects of many different possible emissions pathways (eg), we will only ever see the outcome of one of these. When a model has features like these, assessing its adequacy for purpose is a far more complicated enterprise.

According to Schroeder,[4] the purpose of the IHME model should be understood not as forecasting but as conditional projection. He argues this on the grounds that IHME model documentation includes alongside its estimates caveats like 'assuming social distancing measures are maintained' and excludes arguments to the effect that all US states will (or are likely to) institute such measures (Schroeder, p3).[4] Fuller[3] suggests that it is reasonable to interpret the IHME model results as forecasts (since the model made a single prediction after most states had implemented 'lockdowns') and to interpret the ICL model results as conditional projections (since the model's multiple estimates corresponded to different policy options). At the same time, Fuller[3] questions whether the forecast/unconditional projection distinction is cogent in epidemiology: he notes that epidemic models always make assumptions and wonders whether we should think epidemic models ever make unconditional predictions (see section 3).

To richen this discussion, we should point out that 'projection' is not a homogeneous purpose to which models can be put, nor one that stands opposite 'prediction' in a binary conception of possible model purposes. To be sure, the ICL and IHME models do not have the features of weather models: they do not incorporate precise, local measurements, their predictions are not based on a near-perfect understanding of causal relationships, and the accuracy of their predictions is not easy to test empirically. Rather, the ICL and IHME models have some features in common with climate models: they model coarse-grained statistical summaries of variables (eg, social contact and mortality rates), the models are not high-fidelity representations of their target systems (indeed, the IHME model does not even directly represent infectious disease transmission dynamics), and the accuracy of their predictions is very difficult to test (see section 2.3). However, we should be clear the ICL and IHME models do not make conditional projections in quite the same way as climate models do.

For one, the ICL and IHME models' projections are conditional on policy choices, while climate models are conditional on representative carbon pathways (RCPs). RCPs are not policy choices: they are outcomes that are conditional on policy choices and numerous other factors acting in complex interaction with one another; there are no uncontroversial connections between policy choices and carbon pathways. In comparison, the ICL model (eg) takes policy choices as inputs. Model developers are thus put in the position of estimating how, for example, university and school closures will affect social contact rates—but there is enormous uncertainty around such relationships, not least because they stand to vary from setting to setting. Imagine putting climate modellers in a similar position, for example, asking climate modellers to assume that all people who buy an electric car will receive a US$5000 rebate and to estimate how this will affect the climate. Our confidence in the model would have to be relatively low: the results would no longer reflect a causal pathway of which scientists have a good understanding.

Indeed, the perhaps most crucial difference between the ICL/IHME models and climate models is the way they represent uncertainty. Unlike climate models, which typically include extensive sensitivity analyses and projections are given in terms of possible ranges for 30-year averages,[8] the ICL and IHME models were subject to far more sources of uncertainty than they aimed to account for.[14 15] Despite this, these models provided point projections and narrow confidence intervals that gave the illusion of precision. As Edeling *et al*[14] show, ICL model predictions were merely possibilities within a much wider distribution (the model was highly sensitive to 19 of its 940 parameters)—and, as they say, some measure of uncertainty is required to correctly interpret model results and present a complete picture to policy-makers. Notably, the latter is textbook advice in the field of decision modelling in health economics: to inform health policy decisions, an analytic framework should include a characterisation of uncertainty surrounding the decision, a clear specification of the relevant objectives and constraints, and a means of structuring the decision problem and interpreting model results.[16] (Ch. 6) In our view, distinguishing between models that do and do not follow this advice when aiming to inform policy decisions is just as important as distinguishing between prediction and projection. In other words, we should want to distinguish between projection models that are and are not adequate for the purpose of decision making, which depends in part on how they characterise uncertainty.

Causal models

A separate question that Fuller[3] addresses is whether COVID-19 models should be understood as causal models. This is a difficult question, in part because it is not well defined: causes can be thought of as operating at many different levels of description and a model that looks causal at one level of description might not look so at another. However, Fuller[3] makes a couple of remarks that we should reflect on. The first is that models whose 'structure is derived from the mechanistic theory of how infections spread among individuals' have 'a strong claim to being considered causal models'. The second is that models that give estimates of the effectiveness of policies or behaviour changes in terms like '38 700 000 lives would be saved by a viral suppression strategy' are plausibly being used to make causal inferences. Our view is that a model's incorporation (or not) of mechanistic theory should be irrelevant to us in assessing whether the model is adequate for causal inference, as should be what the model purports to estimate.

To see why, consider a difference between more vs less complex climate models. In complex climate models, a value of 'equilibrium climate sensitivity' (ECS) (a crucial feature of the climate system) is not assumed, but rather depends on the model's estimations of numerous mechanisms (eg, how the heat trapped by the greenhouse-gas-rich atmosphere is transported around and contributes to feedbacks that can trap yet more heat). Less complex models simply put a value of ECS into a model by hand. The more complex models are obviously more 'microcausal' than the less complex ones. However, as long as the ECS value a model assumes is accurate, a less complex model could in principle deliver many of the same answers regarding the causal effect of a carbon pathway as a complex one. Such a simple model might very well be useful, despite the fact that it is not at all derived from the mechanistic theory of how CO_2 leads to warming. In fact, the true mechanism by which CO_2 traps radiation has to do with the way in which a CO_2 molecule vibrates when it is hit by infrared radiation. To our knowledge, no climate model actually represents that causal process.

Similarly, the actual mechanisms by which 'individuals transition from being susceptible to being infectious'[3] involve, among other things, SARS-COV-2 shedding in the upper respiratory tract and the spread of virus-laden respiratory particles, which are influenced by viral load dynamics and patterns of breathing, talking, coughing and airflow, etc.[17 18] No epidemiological model that projects outcomes at the national level represents all of these mechanisms. Even in a comparatively complex individual-based model like the ICL model, whether a susceptible individual becomes infected with SARS-COV-2 at any given time is simply a parameterised function of assumed contact patterns (eg, by age, place) and estimated transmission probabilities per contact (eg, by infectiousness)[11 19]—a function that is not resolved in terms of a mechanistic theory of transmission. This kind of model will be able to make counterfactual projections if and only if the assumed contact patterns and transmission probabilities are correct. And, importantly, if a simulation model is inadequate for counterfactual projection, it will also be inadequate for causal inference. After all, counterfactual projection and causal inference are similar tasks: they are the same type of reasoning applied in different directions.

Ultimately, whether a simulation model is adequate for causal inference depends on many things, but at least one of the following must be true: either the model's results must be insensitive to changes in its parameter values or the correct values for those parameters must be known. Suppose, for example, we want to use a simulation model to infer whether a public health intervention caused a reduction in deaths over the past year. When populated with one set of parameter values, the model correctly retrodicts the number of observed deaths over the past year when (and only when) it assumes the intervention was implemented. When populated with another set of parameter values, the model correctly retrodicts the number of observed deaths over the past year under several assumptions, including the absence of the intervention. Clearly in this case, the model's adequacy for causal inference depends on our knowing which set of parameter values is correct: if we do not know, the model is not adequate for causal inference. We can see, then, why Edeling et al's[14] findings are significant: given that the ICL model was sensitive to specific parameter values (eg, the latent period, the contact rate given social distancing, and the delay to start case isolation) whose correct values were not known, we should not consider the ICL model to be adequate for causal inference. A similar sort of argument can be made about the IHME model, given its reliance on limited mortality curves with known shortcomings.[15] In general, it is the quality and accuracy of the model's mathematical assumptions that represent the effect of A on B that will determine whether the model is good at estimating how a modification to A will affect B— not how fine-grained the model is or at what level of description it looks for causal relationships.

In practice, model users may be tempted to think of models as being more or less fit for causal inference, rather than adequate or inadequate, and to use whatever tools are available to them to perform a pressing task. Nonetheless, some models come with one or more explicit reasons to think they cannot support causal inferences: they lack the degree of fidelity to the real counterfactual relations in the world that is required for that. When modellers present these models to decision-makers, they are taking a risk with foreseeable potential harms, a risk with moral significance.[20]

'Performativity'

Some have argued that some models can have another, distinct function called 'performativity'.[5] Performativity is the capacity of a model to influence the (at least partially social) system that it models. van Basshuysen et al[5] suggest that the ICL model was performative in that it influenced (1) scientific advising, (2) policy decisions and (3) individual behaviour that shaped pandemic outcomes. As van Basshuysen et al[5] note, the first two modes of influence are well documented, while the third is plausible.

One reason for philosophical interest in the concept of model performativity is the apparent challenge it poses for a model's predictive accuracy. As van Basshuysen et al (p 115)[5] note, models like the ICL are designed to provide policy guidance, so their usefulness 'depends on their predictive performance'. At the same time, when a model steers policy-makers away from certain scenarios (eg, a 'do nothing' approach), this makes it difficult to test the accuracy of its projections for those counterfactual scenarios. A good example is the ICL model's projection of 510 000 deaths from COVID-19 in Great Britain and 2.2 million in the USA in 'the (unlikely) absence of any control measures or spontaneous changes in individual behaviour' (Ferguson, p6-7).[11] This particular counterfactual projection is indeed difficult to test. For example, even though we can point to US states that implemented no mitigation measures whatsoever and demonstrate that far fewer deaths occurred there by August, 2020 than the ICL model would have projected[21]

(setting aside here the objection that the ICL model did not directly make state-level projections), it would seem that the model's defenders can always appeal to the possibility that individuals in that state changed their behaviour and undermined the model's projections. Indeed, Schroeder[4] defends the IHME model using a similar tactic: 'although there may be other grounds for criticising the model, we cannot criticise it simply because the death toll has turned out to be far higher than what the model projected. The model told us how many COVID-19 deaths the USA would see if the country implemented strict social distancing measures. Since it did not do that, the actual death toll can't prove the model wrong' (Schroeder, p4). This seems, of course, extraordinarily convenient for model developers. Nonetheless, van Basshuysen et al (p121)[5] worry that when model projections are affected by so-called performativity, this can attract criticisms like the model is 'too pessimistic' and seems to undermine their credentials.

The idea that van Basshuysen et al (p121)[5] explore is that, while a model's performative ability can adversely affect its predictive abilities, this may not imply that a model's 'suitability, adequacy, or usefulness is diminished'. In fact, the authors suggest that we might want, under some conditions, to consider a model's performative impact to be a potential virtue. We have seen similar suggestions pop up in the epidemiology literature: Biggs and Littlejohn (p92),[22] for example, remark that 'Initial projections (of the ICL model) built in worst-case scenarios that would never happen as a means of spurring leadership into action', while Ioannidis et al (p6)[23] speculate that 'In fact, erroneous predictions may have even been useful. A wrong, doomsday prediction may incentivise people towards better personal hygiene'. Assuming for the moment that such claims are true, are there conditions under which inaccurate, pessimistic projections are something to celebrate? Our moral judgement is no: there are no such conditions.

The reason we use decision modelling to inform health policy in the first place is because health interventions have costs, that is, monetary costs, opportunity costs, and/or other undesirable outcomes associated with them, including, sometimes, loss of freedoms that are at the heart of a liberal democracy. The point of models like the ICL model is to help determine if an intervention's benefits warrant those costs. It is only by accurately measuring costs and benefits and knowing what sacrifices we wish to make that we can determine whether performativity (be it defined as changes to scientific advising, policy decisions or individual behaviour) would be desirable or not. Making performativity a goal in model construction, therefore, would be a serious threat to democratic decision making. van Basshuysen et al[5] (p122) seem to appreciate this point, as they concede that it would seem highly problematic to recommend constructing models with certain performative capacities, given concerns around questionable social value influences and the impact on model credentials. Nonetheless, van Basshuysen et al[5] suggest that performativity might be a criterion by which models could be evaluated after the fact. But how could this work? Imagine holding an annual race in which we tell runners that the goal is to complete 10 km in the fastest possible time, but where, year after year, we award the medals to runners who most quickly reach the 5 km mark. Hopefully it is clear that we cannot neatly separate how runners will be evaluated from what they will eventually adopt as their goal. The same will be true of modelling. If those who judge the suitability, adequacy, or usefulness of a model give it high marks when it succeeds performatively (according to the values of the judges in question), they will be sending the signal that modellers should adopt this goal.

Performativity, we argue, is never a legitimate purpose for a model, especially given the risk that the behaviours the model stimulates have costs that people would be unwilling to bear if they had perfect knowledge of them. Consider, for example, that a recent cost-effectiveness analysis of social distancing strategies in Israel estimated that the incremental cost-effectiveness ratio of a national lockdown would be, on average, US\$45 104 156 (median US\$49.6 million) to prevent one death.[24] Set aside the question of whether this figure is correct, or whether this model itself was created with rhetorical purposes in mind. It is clear that saving lives from COVID-19 via lockdowns was at best significantly costly, and reasonable people could disagree over whether the costs, whatever they are, were worth the benefits that they provided. To build an epidemiological model for the purpose of performativity, for example by deliberately producing 'worst-case scenarios', is to stack the deck in favour of certain results of a cost benefit analysis, rather than to perform one. After all, a model that is crafted to stimulate a desired live-saving behaviour is ipso facto one that suggests a comparatively lower cost of that behaviour per life saved. Furthermore, in the absence of a model that makes the highest quality projection, rather than one crafted to be performative, scientists will have no way to know how much the policies they are stimulating will cost—so not only are they deciding on behalf of everyone else without consulting their values, they are doing so blind. We should not rush to defend models that build in 'worst-case scenarios that would never happen' (Biggs and Littlejohn, p92)[22] on the grounds they spur leadership into action: we should pause first and ask what the public values and wants from their leadership.

DUTIES IN MODELLING

In the introduction, we suggested that people involved in scientific modelling have the moral-epistemic duties to establish what purposes a given model aims to serve and to continually assess the model's adequacy for those purposes. Our claim builds on the idea that scientists have a moral responsibility to avoid foreseeable harms,[25] taking into account that using models that are inadequate for purpose can lead to such harms, including the endorsement of false claims and the unjust omission of information,[20] which are significant in the context of decision making. Before we go on, it is important to stress that model purposes come from people, not from models themselves. In principle, any model could be used for almost any purpose—displayed as a work of art, incorporated into one's spiritual practice, or used to make causal inferences or inform policy decisions—regardless of any relevant shortcomings. It is too much to ask for modellers to be responsible for ensuring that their models are never used in a way they do not intend. Rather, it is their duty to establish what they *do* intend and, thus, the scope of their responsibility.

When one is involved in scientific modelling for the purpose of informing decision making, moral and epistemic duties are not as distinct as philosophers often conceive of them[26] but rather cross paths. We call modelling tasks moral-epistemic duties in order to underline that they have moral significance. The moral significance of modelling tasks stems in part from the inherent value-ladenness of modelling, including the representational decisions that are an intrinsic part of the process.[1 20] Representational decisions, in terms of what to represent (eg, in the model structure) and how to represent it (eg, through parameterisation), are uniquely guided by the purpose of the model. Not only is there no such thing as a 'factually correct' way to model something, we often lack anything resembling 'factually correct' model inputs (eg, 'one cannot hope to obtain an

accurate, data-informed value of all parameters in contention' in the ICL model (Edeling *et al* p128)).[14] The result is that representational decisions are routinely informed not by assessments of what is empirically true, but of what is reasonable, sufficient or adequate given the purpose of the model. These are often morally charged decisions, such as whether it is reasonable to exclude certain costs (eg, unintended harms) or adequate to use a certain source of data (eg, from a faraway setting) when estimating the overall benefit of a health intervention.[1] One recognises these decisions as morally significant as soon as one learns that different people would make them differently. Modelling decisions share this in common with all expert judgements: 'If those making judgements share certain characteristics, such as gender, age, race, home ownership, or wealth, they may fail to recognise how the costs of policies (such as stay-at-home orders) are likely to affect those who do not share those characteristics; they might recognise those costs and consider them to some extent but they will not feel those consequences of their decisions' (Moore *et al,* p1).[27]

The fact that modelling decisions are both value-laden and purpose-guided lends an obvious significance to a model's intended purpose; to clarify this purpose is thus a primary moral-epistemic duty in modelling. As we have shown, this task is not as simple as it seems: disambiguating a model's purpose from nearby possibilities requires detailed analysis along multiple axes.[3 4 28 29] Furthermore, a model's purpose is not something that can be easily settled once and for all at the beginning of a modelling project. On the contrary, a model's purpose must continually be re-assessed and understood relative to its epistemic features, both its virtues and shortcomings—in other words, its adequacy-for-purpose. This assessment process involves a complicated weighting of epistemic (scientific) and non-epistemic (social) values that different scientists inevitably carry out differently. This is partly why a question like 'do epidemic models ever make unconditional predictions?'[3] is difficult to answer: epidemiologists with different values may use the same model for different purposes given their different assessments of the model.

Many contributors have documented the predictive failures of certain COVID-19 models,[21–23 30 31] while others have pointed to the (at least relative) success of the same models.[4 19 32] This discrepancy should remind us that there is no universally agreed-upon, objective basis on which to define the accuracy of model predictions. When building a prediction model, modellers' moral-epistemic duties include setting a clear standard for predictive accuracy, not only in terms of precision but in terms of what will count as an accurate prediction. A model's predictions can be accurate for reasons that are not rooted in epistemic virtues—if a model overestimates the impact of a virus on mortality, but also overestimates the efficacy of an intervention, it might end up making an empirically accurate mortality projection in an area where the intervention is adopted— and modellers should be clear about whether or not they value this sort of achievement. At the same time, there is a duty to establish what will count as a failure: often, a model whose purpose is to make conditional projections has an infinite number of built-in 'escape routes', reasons to which the model's defenders can appeal to justify why projections have not panned out in the real world. It is sometimes possible to recognise a model at the outset of development as being unlikely to have strong predictive capabilities, particularly over the long term, given the complexity of its target system, known limitations in the relevant evidence, or other factors. In these cases, the model's purpose should be understood as being to assist decision making under significant uncertainty. This involves special moral considerations, including whether to postpone the decision and pursue additional information.[16]

Beyond a certain degree of uncertainty, utilitarian frameworks for decision making, like those behind harm–benefit and cost-effectiveness modelling, will be of limited use. In such cases, caution is required: we should guard against models being used to justify existing political views by representing their favoured policies as the ones that 'follow the science'. Otherwise, our standards of scientific and democratic scrutiny will suffer.

ORCID iDs
Eric Winsberg http://orcid.org/0000-0002-8764-0072
Stephanie Harvard http://orcid.org/0000-0002-9008-2163

REFERENCES

1. Harvard S, Winsberg E, Symons J, *et al*. Value judgments in a COVID-19 vaccination model: a case study in the need for public involvement in health-oriented modelling. *Soc Sci Med* 2021;286:114323.
2. Winsberg E, Brennan J, Surprenant CW. How government leaders violated their epistemic duties during the SARS-CoV-2 crisis. *Kennedy Inst Ethics J* 2020;30:215–42.
3. Fuller J. What are the COVID-19 models modeling (philosophically speaking)? *Hist Philos Life Sci* 2021;43:47.
4. Schroeder SA. How to Interpret Covid-19 Predictions: Reassessing the IHME's Model. *Philosophy of Medicine* 2021;2.
5. Van Basshuysen P, White L, Khosrowi D, *et al*. Three ways in which pandemic models may perform a pandemic. *Erasmus Journal for Philosophy and Economics* 2021;14:10–127.
6. Harvard S, Winsberg E. Causal inference, moral intuition, and modeling in a pandemic. *Philosophy of Medicine* 2021;2.
7. Harvard S, Werker GR, Silva DS. Social, ethical, and other value judgments in health economics modelling. *Soc Sci Med* 2020;253:112975–9.
8. Winsberg E. *Philosophy and climate science*. Cambridge: Cambridge University Press, 2018.
9. Süsser D, Ceglarz A, Gaschnig H, *et al*. Model-Based policymaking or policy-based modelling? how energy models and energy policy interact. *Energy Research & Social Science* 2021;75:101984.
10. Airoldi G, Vecchi D. The road from evidence to policies and the erosion of the Standards of Democratic scrutiny in the COVID-19 pandemic. *Hist Philos Life Sci* 2021;43:66.
11. Ferguson NM, Laydon D, Nedjati-Gilani G. Report 9: impact of Non-Pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID response team, 2020. Available: https://wwwimpe-rialacuk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/
12. Murray CJL, Institute for Health Metrics and Evaluation Covid 19 Health Service Utilization Forecasting Team,. Forecasting Covid 19 impact on hospital bed days, ICU dA Ys, ventilator days and deaths by US state in the next 4 months. *MedRxiv preprint* 2020.
13. Vemer P, Krabbe PFM, Feenstra TL, *et al*. Improving model validation in health technology assessment: comments on guidelines of the ISPOR-SMDM modeling good research practices Task force. *Value in Health* 2013;16:1106–7.
14. Edeling W, Arabnejad H, Sinclair R, *et al*. The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nature Computational Science* 2021;1:128–35.
15. 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;173:226–7.

16 Briggs A, Sculpher M, Claxton K. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press, 2015.

17 Cevik M, Tate M, Lloyd O, *et al*. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe* 2021;2:e13–22.

18 Jayaweera M, Perera H, Gunawardana B, *et al*. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. *Environ Res* 2020;188:109819.

19 Rice K, Wynne B, Martin V, *et al*. Effect of school closures on mortality from coronavirus disease 2019: old and new predictions. *BMJ* 2020;371:m3588.

20 Harvard S, Winsberg E. The epistemic risk in representation. forthcoming in Kennedy insT. *Ethics J. Pre-print* 2021 http://philsci-archivepittedu/18576/

21 Winsberg E, Brennan J, Surprenant C. This paper attacks a Strawman but the Strawman wins: a reply to van Basshuysen and white. *Kennedy Inst Ethics J* 2021;31:429–46.

22 Biggs AT, Littlejohn LF. Revisiting the initial COVID-19 pandemic projections. *The Lancet Microbe* 2021;2:e91–2.

23 Ioannidis JPA, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. *Int J Forecast* 2020;382.

24 Shlomai A, Leshno A, Sklan EH, *et al*. Modeling social distancing strategies to prevent SARS-CoV-2 spread in Israel: a cost-effectiveness analysis. *Value in Health* 2021;24:607–14.

25 Douglas H. *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press, 2009.

26 Feldman R. Epistemological Duties. In: Moser PK, ed. *The Oxford Handbook of Epistemology*. New York: Oxford University Press, 2005.

27 Moore A, MacKenzie MK. Policy making during crises: how diversity and disagreement can help manage the politics of expert advice. *BMJ* 2020;26:m4039.

28 Tufekci Z. Don't believe the COVID 19 models. The Atlantic, 2020, 2020. Available: https://www.theatlantic.com/technology/archive/2020/04/coronavirus-models-arent-supposed-be-right/609271/

29 Fuller J. Conceptual lessons for containing an epidemic of questionable COVID-19 counterfactual comparisons. *J Epidemiol Community Health* 2021;75:1031–2.

30 Chin V, Samia NI, Marchant R, *et al*. A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York state. *Eur J Epidemiol* 2020;35:733–42.

31 Marchant R, Samia NI, Rosen O. Learning as we go: an examination of the statistical accuracy of covid19 daily death count predictions, 2020. Available: https://arxiv.org/abs/2004.04734

32 van Basshuysen P, White L. Were lockdowns justified? A return to the facts and evidence. *Kennedy Inst Ethics J* 2021;31:405–28.