Received: 26 May 2020 | Revised: 7 July 2020 | Accepted: 19 July 2020

DOI: 10.1111/jep.13459

ORIGINAL PAPER

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

WILEY

Agent-based modelling for SARS-CoV-2 epidemic prediction and intervention assessment: A methodological appraisal

Mariusz Maziarz BSc, MA[1,2] | Martin Zach MA[3]

[1]Interdisciplinary Centre for Ethics, Jagiellonian University, Kraków, Poland

[2]Institute of Philosophy, Jagiellonian University, Kraków, Poland

[3]Department of Philosophy and Religious Studies, Faculty of Arts, Charles University in Prague, Prague, Czech Republic

Correspondence
Mariusz Maziarz, Interdisciplinary Centre for Ethics and Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland.
Email: mariusz.maziarz@uj.edu.pl

Funding information
Grantová Agentura České Republiky, Grant/Award Number: GA ČR 19-04236S; European Research Council, European Union's Horizon 2020, Grant/Award Number: 805498

Abstract

Background: Our purpose is to assess epidemiological agent-based models—or ABMs—of the SARS-CoV-2 pandemic methodologically. The rapid spread of the outbreak requires fast-paced decision-making regarding mitigation measures. However, the evidence for the efficacy of non-pharmaceutical interventions such as imposed social distancing and school or workplace closures is scarce: few observational studies use quasi-experimental research designs, and conducting randomized controlled trials seems infeasible. Additionally, evidence from the previous coronavirus outbreaks of SARS and MERS lacks external validity, given the significant differences in contagiousness of those pathogens relative to SARS-CoV-2. To address the pressing policy questions that have emerged as a result of COVID-19, epidemiologists have produced numerous models that range from simple compartmental models to highly advanced agent-based models. These models have been criticized for involving simplifications and lacking empirical support for their assumptions.

Methods: To address these voices and methodologically appraise epidemiological ABMs, we consider AceMod (the model of the COVID-19 epidemic in Australia) as a case study of the modelling practice.

Results: Our example shows that, although epidemiological ABMs involve simplifications of various sorts, the key characteristics of social interactions and the spread of SARS-CoV-2 are represented sufficiently accurately. This is the case because these modellers treat empirical results as inputs for constructing modelling assumptions and rules that the agents follow; and they use calibration to assert the adequacy to benchmark variables.

Conclusions: Given this, we claim that the best epidemiological ABMs are models of actual mechanisms and deliver both mechanistic and difference-making evidence. Consequently, they may also adequately describe the effects of possible interventions. Finally, we discuss the limitations of ABMs and put forward policy recommendations.

KEYWORDS

agent-based modelling, causal inference, difference-making evidence, mechanism, mechanistic evidence, SARS-CoV-2

1 | INTRODUCTION

In the aftermath of the outbreak of the novel coronavirus, governments around the globe have introduced non-pharmaceutical public health interventions aimed at slowing down the spread of the resultant pandemic. These measures range from relatively mild requirements like wearing face masks, washing hands, or avoiding close contacts to school closures and imposed isolation that are likely to have a detrimental and unpredictable influence on social and economic life.[1] Despite their significant impact, the introduction of many of these measures was not supported with high-quality evidence. First, conducting RCT would not be feasible for both ethical and practical constraints. Second, significant differences between the coronaviruses that caused the SARS and MERS outbreaks and SARS-CoV-2 (such as the likely airborne transmission[2] and asymptomatic infectiousness of the latter[3,4]) undermine extrapolation from the data gathered during these previous epidemics. Finally, the current pandemic has not lasted long enough to gather observational data in the amount and quality sufficient for the assessment of the efficacy of alternative public health interventions, since the first reports were published just weeks after the first measures were introduced.[5]

One of the many ways to address the issue concerning the impracticality of conducting RCTs and observational studies in the context of an ongoing pandemic is through scientific modelling, in particular epidemiological modelling. Here, we focus on the so-called agent-based modelling (ABM) approach, which differs from more traditional epidemiological modelling in several ways.

ABMs are a form of computational modelling strategy where agents are treated as entities interacting with each other and their environment in a locally defined fashion described by a set of rules. The overall dynamics of the system are then computed, allowing for the simulation of complex patterns and an understanding of how these patterns arise.[6,7] ABMs are used in many scientific contexts, including modelling the spread of infectious diseases, and have proven successful in informing policy decisions before. For instance, Eisinger and Thulke[8] modified and then applied a previously developed ABM of the spread of rabies, generating a rule-based model that represented specific spatial and behavioural characteristics of the fox population (eg, with fox families represented as moving within home ranges and young foxes engaging in long-distance migratory behaviour).[6] Whereas the classical differential equation models predicted that vaccinating at least 70% of the fox population would eliminate rabies, the ABM indicated that a successful vaccination strategy could do with much less than 70% of the population being immunized once the spatial arrangements of fox hosts were explicitly considered, saving millions of Euros as a result. Moreover, the ABM also suggested that the classical strategy would fail more often than not, and was successfully applied to deal with the rabies problem. However, despite the promising record of using ABMs in effective epidemiological interventions, its use in informing proposed measures against the novel coronavirus epidemic has raised criticism.[9-11]

Unfortunately for the assessment of healthcare interventions based on this type of epidemiological models, standard evidence hierarchies exclude agent-based models altogether and include theoretical or mechanistic inferences at the lowest level of the hierarchy. For example, the Oxford Centre for Evidence-Based Medicine[12] and the National Institute for Health and Care Excellence (NICE guidelines)[13] include theoretical and mechanistic reasoning but agent-based models fall beyond their scope. This can be explained by the novelty of agent-based modelling and the limited trust of EBMers in theoretical and, to some extent, also mechanistic reasoning, which, despite being used implicitly to assess the possibility of confounding and the quality of results,[14] is downgraded or rejected as either subjective or fallacious.[15] However, such a view has been challenged by a group of philosophers advocating for improving the practices of evidence assessment in medicine by putting more weight on mechanistic reasoning in causal inference.[16-18] The position of the EBM+ program[16-18] is encapsulated by the normative reading of the Russo-Williamson Thesis,[19] which states that causal claims should be based on both difference-making and mechanistic evidence.

The causal claims supported by agent-based models have been interpreted inconsonantly: either as being in line with the potential outcome approach (POA),[20] as delivering theory-driven understanding[21] or as providing mechanistic evidence.[22] Below, we show that all of these apparently inconsistent interpretations are correct, because the best contemporary ABMs bear a resemblance to the actual mechanisms and therefore allow for the counterfactual assessment of intervention efficacy in the target while also delivering an understanding of the phenomena of interest. Our argument proceeds by (a) discussing as a case study an ABM of SARS-CoV-2 epidemic in Australia, (b) showing that the best ABMs represent actual mechanisms despite the presence of various simplifications and (c) considering the limitations of using ABMs as evidence for clinical and policy decisions.

2 | MODELLING THE SARS-COV-2 EPIDEMIC

Apart from the compartmental SIR (Susceptible, Infectious, Recovered) framework and its derivatives[23-28] or regression analysis,[29,30] most advanced models of the spread of the novel coronavirus are transformed versions of agent-based influenza pandemic models.[11,31] Such models have been used as evidence for introducing (sometimes severe) public health measures,[32] with the recent change in British policy being the prime example. In this section, we illustrate this approach to modelling the SARS-CoV-2 pandemic with an agent-based model of the epidemic in Australia[31] based on AceMod. Developed as a "framework for studying influenza pandemics in Australia"[33] (p. 412). AceMod is an influenza spread model that addresses the need for simulating interventions responding to the outbreaks of future respiratory diseases. While the 2009 swine flu pandemic was the motivation for constructing AceMod, the model was not intended to accurately represent the outbreak of the H1N1 strain, but rather as a generalized framework for studying how an infectious disease spreads through the social interactions of

Australians. AceMod utilizes census data to ascribe realistic spatial and social characteristics to almost 20 million agents inhabiting the model world. These agents are divided into different social groups of varying characteristics, with households differentiated proportionally according to statistical data on the prevalence of different types of families (singles, single parents and couples with or without children). These features are ascribed to agents stochastically in a way that replicates the aggregate structure of statistical data. During the daytime, children and students meet in classrooms and at schools, adults go to work and pensioners stay at home. During the nighttime, the agents encounter contacts at households and in their neighbourhoods (eg, at supermarkets, theatres).

The disease can be contracted by an agent in the event of meeting an infected individual in one of these settings. The probability that an agent *i* contracts the disease in a given step *t* depends on the number of sick individuals met in that step and the contagiousness of the disease, scaled by *K*. The modellers assume that the infectivity of the disease decreases linearly over time. Asymptomatic cases are assumed to be 50% less infectious than symptomatic ones, and the flu lasts 5 days within the model. After this period, recovered agents cannot infect others. Additionally, those who experience symptoms do so after an incubation period lasting approximately 3 days. The influenza epidemic is started by agents coming to Australia via international airports and seeded into communities living near the airports at random.

To represent an epidemic of a particular strain of influenza with AceMod, the model requires calibration. Modellers can proceed with this step in two ways, depending on the accessibility of data. In the case of well-studied influenza strains, their infectivity and the ratios of transmission in different contexts are well-recognized, and parameter values can be chosen based on empirical studies. However, if these data are missing, then parameter values have to be calibrated using statistical procedures such as simplex or genetic algorithms to maximize the fit of the model to a benchmark. After constructing and calibrating AceMod, modellers run simulations to obtain the estimates of prevalence, incidence and attack rates, and choose the most common outcome (due to stochasticity, different runs of the model may lead to obtaining slightly different results).

Chang et al[31] have used a significantly amended version of AceMod to address the question of the effectiveness of non-pharmaceutical interventions aimed at suppressing the SARS-CoV-2 epidemic in Australia. The selection of models constructed to control a novel and possibly deadly strain of the seasonal flu in this case is primarily the result of the rapid demand for evidence informing decisions regarding public health measures, which may raise doubts about the justification and soundness of their conclusions. For example, one can ask whether the efficacy claims assess healthcare interventions against the novel coronavirus epidemic or an artificial pathogen existing only within the model world that shares some features of influenza and others of SARS-CoV-2. To address this criticism (considered in depth below), we discuss the changes introduced to the model and argue that the process of model calibration and validation suggests that the model represents the actual mechanism of the SARS-CoV-2 epidemic.

ABMs such as AceMod can be seen as consisting of two parts: the rules specifying the behaviour of agents and the creation of the model society, as well as the assumptions characterizing the infectivity of the pathogen causing the epidemic. Given that AceMod is based on 2016 census data and a major change in social behaviours is unlikely to have occurred since then, the model accurately represents the social interactions of present-day Australians. Hence, the former part of the model has been left mostly unchanged, beyond increasing the number of agents to over 24 million to adjust for the growing population. In addition to introducing a social structure sufficiently resembling the contact network of the present population, obtaining accurate predictions of epidemic development and policy assessment requires inputting data on transmission likelihoods that are true for the pathogen causing the modelled epidemic.[34] Most changes in the model are concerned with the assumptions specifying the infectivity of the disease. Even though several features of influenza epidemics are similar to the epidemic caused by the novel coronavirus, they differ with respect to infectivity and attack rates, mortality rates, the average duration of disease, the reproductive number R_0 and the distribution of asymptomatic cases. Therefore, these parameters in the model required recalibration.

The transmission probabilities remained mainly as specified in the influenza model. To account for the differences in the incubation period and disease length, Chang et al set the time from contraction to the appearance of symptoms to 5 days on average and the duration of the disease to 12 days. Infectivity increases exponentially the day after an agent gets infected and then decreases linearly until the end of infection, so cases are most infectious at the start of symptoms. The length of the generation period was calibrated to 6.4 days to reflect this difference in the model. Additionally, the likelihood of contracting SARS-CoV-2 but staying asymptomatic was set to be age-dependent, and equalled 1/3 for adults while minors were set to be five times less likely to suffer from symptoms than adults. While this assumption is in agreement with the empirical findings that children represent a minor fraction of symptomatic cases, the calibration aimed at reproducing aggregate epidemic curves and may diverge from the actual chances of developing symptoms.

Within the AceMod framework, the reproductive number R_0 is not one of the assumptions inputted into the model. Rather, its estimate results from a simulation of the scenario described by the rules and assumptions, some of which are stochastic. The assumptions considered and, particularly, the parameter denoting contagiousness of the disease (*K*) have been calibrated such that R_0 stays within the limit of (2.0-2.5), that is, in agreement with empirical estimates of the reproductive number at the beginning of the SARS-CoV-2 outbreak.[35,36] The set of parameter values that result in the estimate of $R_0 = 2.27$ create the epidemic dynamics reproducing the beginnings of the outbreak in a few countries experiencing the disease prior to Australia (China, Italy, Spain), where the growth rate of cumulative incidence equalled roughly 0.2. In addition to reproducing the empirical data for the beginning of the epidemic, the recalibrated AceMod allows for simulating what the future of the epidemic in Australia may look like. As the modellers admit, the *Baseline* scenario, which is based

on the assumption that agents do not change their behaviour in response to the epidemic, is unlikely given the widespread self-imposed isolation in other countries. However, it allows for counterfactual comparisons of the different possible (sets of) interventions relative to the Baseline scenario. To assess the efficacy of particular healthcare policies, Chang et al modify relevant rules and assumptions to describe the spread of SARS-CoV-2 under either case isolation, school closure, along with three levels of compliance with social distancing, along with a few combinations of the three policies. For instance, to assess the effect of school closure (including primary and secondary schools, colleges and universities), the parameter denoting the chance of meeting an infected agent in schools is set to zero, which describes the situation when both students and teachers stay at home (and hence cannot contract the virus). These counterfactual scenarios represent the effects of interventions on the model world. All interventions are modelled as taking place after the number of cases exceeds 1000. The comparison of most common outcomes (given the stochasticity of the assumptions and rules, they are indeterministic) including interventions with the baseline scenario allows for putting forward counterfactual causal claims that describe the effects of interventions on peak incidence and prevalence and the development of the epidemic in time. The conclusions accurately describe the effects of interventions within the model as long as no coding error occurs. However, the reliance of the model on simplifications generates a question as to whether the assessment of intervention efficacy holds for the novel coronavirus epidemic in Australia.

3 | ABMS AS MODELS OF ACTUAL MECHANISMS

Before proceeding to our argument, let us first make several general remarks about modelling. These remarks should prove essential in clarifying the main issues that are often raised with regard to using simplified models, particularly in the context of policy decision-making. First of all, ABMs are instances of mechanistic models, for they clearly fit the general, also called the minimal, characterization of what a mechanism is: a set of entities whose activities and interactions are organized such that they are responsible for the phenomenon.[37-39] This definition is broad enough to conceptually unify the debates on biological and social mechanisms under a single notion of a mechanism. Furthermore, such definition leaves open the possibility of integrating biological and social aspects into a mixed-mechanism model.[40]

It should also be noted that much like any other kind of model, ABMs serve as simplified representations of their target phenomena. As the AceMod case clearly shows, modellers introduce various simplifications by which they purport to adequately capture the core dynamics of the modelled phenomenon. In this process, they first abstract away from the complexities of the real system by "extracting" certain features that they believe to be of crucial importance and that will then be the focus of modelling, whereas other features that may or may not have a causal influence are disregarded in these early stages. Modelling is an iterative process during which the merits of the model's assumptions are continuously being evaluated, and if required, the assumptions are refined and additional assumptions added. More importantly, some of those extracted features are distorted to the extent that, if taken literally, they would misrepresent the actual state of things. However, introducing such distortions is often made in full awareness, with the ultimate goal of finding out whether the consequences they have for the behaviour of the system make a difference and to what degree. Philosophers often refer to the former—that is, the set of properties retained in a model—as an abstraction, while the latter case—that is, the distortions of the system's features—is called an idealization.[41]

However, abstractions and idealizations do not exhaust the conceptual toolbox available to modellers. A popular way to attempt to model a given system realistically is to introduce various approximations. Although there are noteworthy differences between approximations and idealizations, we cannot afford to go into any detail here. In summary, models often effectively disregard, distort and otherwise simplify possibly important details. In light of this, many wonder whether we can gain insight into the modelled phenomenon at all, and if so then how.

Although the SARS-CoV-2 ABM is fairly detailed and precise, it cannot do without some of the simplifications discussed above. Consider some of the following assumptions introduced in the model. On the one hand, the basic features of the social life of the majority of the population are extracted and considered in the model: for example, the inclusion of day and night regimes with their respective differences in social behaviour allows for modelling a more realistic scenario than in simpler models. On the other hand, the infectivity of symptomatic and asymptomatic cases is considered to be constant for all members of the two groups of agents, albeit it differs between the groups. In reality, we expect that infectivity varies, which is further supported by extreme cases of super-spreaders who infect a large number of people and thus may seed new local outbreaks, which could arguably impact the predictions.[42-44] Other parameter values also have a wide distribution but are treated as constant, often by calculating the mean value. The ABM also does not consider the potential impact of ethnic differences[45-49] in the population with respect to differing lifestyles, socioeconomical status and immune host responses, all of which could affect the dynamics of the spread. Unfortunately, Chang et al[31] have not conducted sensitivity analysis and therefore we lack evidence for assessing the influence of such simplifications on model predictions.

Furthermore, some other assumptions exceed our current understanding of the epidemic and SARS-CoV-2's transmission mechanism. For example, one of the assumptions of the AceMod model is the linear reduction of infectivity over time. Unfortunately, empirical results[50] suggest only that infectivity reduces over time, but do not indicate the linearity of this process. Additionally, AceMod and its SARS-CoV-2 model put agents into working groups of 20 agents, despite the heterogeneity of their working conditions. Considering the differentiation of work duties (from healthcare workers and shop assistants to writers with virtually no social interaction), the chance of

meeting an infected person at work is actually job-specific and therefore the model simplifies the reality.

Consequently, we concur with Andersen's claim that "no mechanism model can include all the actual, much less the potential, causal relationships in which such a mechanism may engage in a system"[51] (p. 995). This pessimistic view on simplified models has inspired the method known as exploratory modelling.[52] In cases when the values of parameters and assumptions inputted into the model cannot be established with certainty, researchers can simulate multiple possible worlds to discover the dependencies that are stable across the set of different models. In cases when only a fraction of assumptions are uncertain, researchers conduct sensitivity analyses to check if changes in the values of the parameters lead to changes in their conclusions.[53] The results that remain unchanged despite minor adjustments to assumptions are considered to be robust.[54] This, in turn, leads to choosing those interventions that are most effective across different sets of parameter values, known as robust decision-making.[52]

Others prefer to think in terms of the distinction between how-actually and how-possibly modelling, referring to models that describe an actual mechanism or a possible mechanism, respectively.[55] There are two general ways to unpack the concept of a how-possibly model. First, we may want to say that a model serves as a hypothesis to be confirmed or disconfirmed as new evidence emerges. In this sense, a how-possibly model will eventually either turn into a how-actually model, should the evidence confirm it, or be discarded if the evidence is contrary to the model's conclusions. The other general notion of a how-possibly model invites a different attitude. Rather than being in the position of having little data to establish whether or not the model does, in fact, represent the actual mechanism, we may interpret the model as representing something other than the potentially actual mechanism. On this view, claims about possible mechanisms do not attempt to pick out actual states, nor do they attempt to explain how a phenomenon actually occurs. Instead, they refer to conceivable states, and ask whether the hypothesized mechanism could, in principle, produce the phenomenon in question if certain assumptions are satisfied.

Here we argue that, notwithstanding the simplifications introduced in the discussed influenza and SARS-CoV-2 ABMs, the epidemiologists are, in fact, providing representations of actual mechanisms of the spread of the viruses. This can be supported by exploiting the relevant similarities[56,57] between the SARS-CoV-2 ABM and the actual outbreak. The respects in which an ABM can be judged similar to its target concern the features retained in that model, while the degree(s) of similarity concern the extent to which the model's features match those of the phenomenon. A good example is setting the parameter/assumption of incubation period = 5 days. This assumption was introduced based on empirical research: "We maintained the incubation period (the interval from the infection to the onset of disease in an individual) around the mean value of 5.0 days, as reported in several studies, for example, the mean incubation period was reported as 5.2 days, 95% confidence interval (CI), 4.1 to 7.0, while being distributed around a mean of approximately 5 days within the range of 2 to 14 days with 95% CI"[31] (p. 3).

To elaborate this further, Glennan[58] introduced a useful conceptual distinction between what he called behavioural adequacy and mechanical adequacy. According to Glennan, a model represents an actual mechanism if it reproduces the aggregate behaviour of the phenomenon, and truthfully describes its parts and interactions. Concerning the behavioural adequacy, one should be asking if "the model predict[s] (quantitatively or qualitatively) the overall behaviour of the mechanism?"[58] (p. 457). By calibrating the model to data from the beginning of the epidemic, Chang et al[31] showed that it reproduces the benchmark variables (R_0 and attack rate).

Two remarks are in order here. First, one may oppose the claim that what is being represented is the actual mechanism by arguing that the mechanism underlying the beginning of the outbreak and the fully-fledged epidemic are distinct. Changes in social behaviour or genetic mutations could undermine the behavioural adequacy of the model. Second, it is possible (at least in principle) that the model represents a false mechanism, but is calibrated to the relevant benchmark such that it reproduces it. For example, there is no data confirming (or disproving) the assumption that children are asymptomatic five times more often than adults. As the modellers admit, this assumption was made not only to account for the lower attack rate among minors, but also to make the model adequate to aggregate-level data. This approach to calibration resembles the estimation of statistical parameters (a.k.a. curve fitting) and is considered dubious. The main line of criticism highlights that it is in principle possible to construct a model that represents a possible mechanism and, using calibration, adjust parameter values so that it reproduces the represented phenomenon, that is, obtains behavioural adequacy despite being false. However, while this criticism is indeed justified regarding models of mechanisms that are epistemically inaccessible in other ways (such as mechanisms in the social sciences[59]), it is not so in the case of epidemiological mechanisms whose transmission mechanism can be studied empirically and compared to the mechanism represented by the model.

This can establish that the mechanism represented by the model is similar (in relevant aspects and to relevant degrees) to the mechanism that generates the outbreak, that is, achieves mechanical adequacy in Glennan's terminology. Applying the list of Glennan's[58] (p. 457) criteria for mechanical adequacy justifies the claim that the mechanism represented by Chang et al[31] resembles the actual mechanism. First, according to our best contemporary understanding of the spread of the novel coronavirus, the model identifies all of the components of the mechanism. This would change if further studies identified other significant transmission routes, for example, the faecal-oral route. Second, the model represents the entities of the mechanism in a localized way, given that it retains the spatial distribution of inhabitation in Australia. Additionally, the model simulates the development of an epidemic in time. This asserts that the "spatial and temporal organization of the mechanism" is accurately represented. Third, given that the number and place of social interactions are crucial for modelling the spread of contagious diseases, the model accurately captures relevant properties of the agents inhabiting the model world. Fourth, the calibration to census data asserts that the model provides "quantitatively accurate descriptions of the interactions and activities of each

component," at least on average for groups of agents. Finally, our background knowledge suggests that there is no other mechanism (different from the spread of the pathogen through human interactions) that could be responsible for the epidemic of SARS-CoV-2.

Given that AceMod fulfils Glennan's criteria for behavioural and mechanical adequacy, considering our current understanding of the novel coronavirus, we can conclude that Chang's et al[31] model represents the actual mechanism of the spread of the disease in Australia. Given this, the claims assessing the efficacy of the mitigation measures under consideration are likely to be accurate not only within the model but also about its target. We claim this with several caveats in mind to be discussed in the next section.

It is also important to note that the ABM integrates the biological aspects, expressed by the parameter of infectivity, and the social aspects such as daily interaction regimes. As a result, the ABM should be construed as an instance of a model of a mixed mechanism, a concept elaborated by Kelly et al.[40] Due to exposure patterns, population-level phenomena such as infectious disease epidemics are crucially dependent on human behaviour and social practices. In cases like the current pandemic, effective interventions may best be aimed at the societal level and therefore mechanistic models that integrate social factors, human behaviour and biological aspects (something that the ABM discussed here attempts to do) are arguably best suited for providing understanding and suggesting policy decisions.

4 | DISCUSSION AND RECOMMENDATIONS

Our study defends using ABMs for informing decisions regarding mitigation and suppression measures by arguing that its best epidemiological models represent actual mechanisms. Provided that the model's assumptions are calibrated and checked against the background empirical data—that is, the components, their activities, and spatio-temporal organization resemble (in relevant aspects and to a certain degree) the actual state of things—iterative runs of the simulations can indeed provide understanding and inform policy decisions. This is because the model delivers both difference-making and mechanistic evidence by satisfying the criteria of behavioural and mechanical adequacy, respectively.

In contrast to our claim, epidemiological SIR models and ABMs have been criticized for over-simplifying target phenomena and hence lacking relevance for policy decisions. For instance, Eubank et al criticized the Imperial College London model[11] for its "reliance on a simplified picture of social interactions [that] limits its extensibility to counterfactuals. The general nature of conclusions based on such model can be expected to be similar to those of a simple compartmental model"[60](pp. 5-6). Similarly, Squazzoni et al suggested that even though AceMod is better calibrated than other epidemiological ABMs, "these [models] do not capture network effects nor people's reactive responses as the population states simply change via stochastic (randomized) processes determined by parameters (although the parameters derive from data)"[9](p. 2.6). In our view, these highly advanced epidemiological models, while being simplified representations of reality, account for relevant aspects of social interactions and crucial aspects of the novel coronavirus epidemic (eg, contagiousness), therefore allowing them to be put forward as evidence for policy-relevant claims.

We claim this despite a straightforward comparison of model predictions to the actual epidemic curve (eg, the number of total cases) in Australia shows the two to be mismatched. The number of covid-19 cases is smaller than predicted by an order of magnitude. However, such a direct comparison is not warranted because the countermeasures implemented by the National Cabinet and the state governments differ from the mitigation and suppression interventions considered by Chang et al.[31] That is, the a posteriori behavioural adequacy of the model cannot be directly assessed based on the predictions because the scenarios implemented into the model differ from the actual course of events. In particular, first restrictions on international travel were imposed on March first, when just 29 COVID-19 cases were observed,[61] followed by the 14-day quarantine for incomers[62] on 15th March (300 cases)[61] that virtually stopped the import of new cases to Australia, the closure of borders for nonresidents[63] and a social distancing rule (requiring 4 m^2 for each person in enclosed space)[64] on 20th March (928 cases).[61] Two days later (1609 cases),[61] some states closed non-essential businesses[65] and, on 30th March (4460 cases),[61] forbade gatherings of more than two people and advised staying at home with some exceptions.[66] The last two interventions are more severe than the measures considered by the modellers and are a plausible explanation of the overestimation of the number of cases. Given this, we can claim that the model had been behaviourally adequate to the mechanism governing the beginning of the epidemic in Australia and it would produce accurate predictions if the interventions were introduced in line with the measures simulated by Chang et al.[31] However, inaccurate predictions are what should be expected in the case of the so-called fat-tail processes, where outcomes strongly depend on the initial conditions. One should expect that, over time, the assumptions and calibrated parameters will be more accurate and ABMs will produce predictions not only qualitatively but also quantitatively accurate. The usefulness of epidemiological ABMs for decision-makers results from delivering an understanding of the spread of the virus and allowing for comparisons among alternative mitigation measures. For instance, one of the qualitative predictions of the model is the limited efficacy of school closures, which remained open in Australia[67] and had limited influence on the severity of the epidemic, considering that just one cluster was located at a school.[68]

We believe that, considering the diversity in the number and patterns of social interactions across countries, the quality of evidence from ABMs should be assessed on the case by case basis. To do so, one can employ the approach of Parkinnen et al[17](p. 79) developed initially to evaluate the quality of evidence for biological mechanisms. In that case, one should consider (a) the quality of the method (ie, consider the empirical adequacy of the assumptions in light of contemporary empirical results), (b) the implementation of the method (ie, assess how the epidemiological ABM is programmed, calibrated and simulated) and (c) the stability of the results (ie, how sensitive the results are to changes in the assumptions). AceMod[31] fulfils the first

two criteria (provisionally accepting the existing empirical results but keeping in mind that they may change as the pandemic develops in time and new results become published), and assessing the third one is impossible with the publicly available data.

Epidemiological models usually do not account for the harms of non-pharmaceutical interventions. Severe mitigation measures such as imposed social distancing and business closures are likely to hamper economic and social life. All models are partial representations of reality and, given that the primary purpose of an epidemiological model is to address the efficacy of health care interventions, they isolate away certain factors and effects of interventions (economic and social) and are more accurate in predicting the spread of the disease under alternative conditions. Other models[69,70] trade-off epidemiological accuracy with accounting for social and economic effects, and may be more relevant for assessing the harms of mitigation measures.

Additionally, ABMs, much like the compartmental models, are dependent on the assumptions of the modellers.[10] Our claim that AceMod calibrated for SARS-CoV-2 bears similarity to the actual mechanism of the epidemic depends on the accuracy of the empirical results used as an input for this model. We need to repeatedly acknowledge the provisional nature of these empirical results, given the novelty of the pathogen. If the parameter values in AceMod were miscalibrated, then the assessments of intervention efficacy could be wrong. This implies that neither the virus can mutate nor that people can significantly and unpredictably change their behaviour since "the efficacy of implementation depends on people's reactions, [the stability of] pre-existing social norms and structural societal constraints."[9] Furthermore, the effects of epidemiological agent-based modelling are highly dependent on social structure and carefully calibrated to social and economic characteristics. Therefore, the epidemiological ABMs are geographically localized and their conclusions should not be extrapolated beyond their target systems,[71] unless the models and their predictions are calibrated to particular settings. Finally, while AceMod is well-documented in the two publications discussed throughout our paper, neither its code nor detailed documentation regarding its use is published (this unfortunately also applies to some other ABMs of the SARS-CoV-2 epidemic). Given these limitations, the models should be carefully checked for coding errors and other possible flaws before applying their implications in the policy context.

In summary, we have argued that, despite the criticism raised against models being the appropriate vehicle for informing policies, the SARS-CoV ABM is suitable for this purpose because the mechanism described by the model sufficiently resembles the mechanism at work in the real world. Thus, our best contemporary epidemiological ABMs are representations of the actual mechanism of the spread of the virus. Unfortunately, such models have been left out from methodological discussions and are not explicitly listed by evidence hierarchies. While the need for appraising mechanistic reasoning in medicine is also voiced by EBMers,[72] there is no broadly-accepted view on how to amalgamate evidence of different types. Further research is needed to assess the risk of bias in the epidemiological models that deliver both difference-making and mechanistic evidence. However, considering the current situation and pressing need for rapid and accurate decisions regarding mitigation measures, policymakers should take to heart the advice that "if no randomized trial has been carried out [...], we must follow the trail to the next best external evidence and work from there"[73] (p. 74). In the current situation, accurately calibrated epidemiological ABMs are the best existing evidence.

ACKNOWLEDGEMENTS

The work of Mariusz Maziarz has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 805498). Mariusz Maziarz recieved scholarship from the Foundation for Polish Science (FNP). The work of Martin Zach was supported by the Czech Science Foundation, project GA ČR 19-04236S.

CONFLICT OF INTEREST

None of the authors reports conflict of interests.

ORCID

Mariusz Maziarz https://orcid.org/0000-0003-1979-0746
Martin Zach https://orcid.org/0000-0001-7181-0391

REFERENCES

1. Wilder-Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *J Travel Med*. 2020;27(2):taaa020. https://doi.org/10.1093/jtm/taaa020.
2. Lewis D. Is the coronavirus airborne? Experts can't agree. *Nature*. 2020;580(7802):175-175. https://doi.org/10.1038/d41586-020-00974-w.
3. Li C, Ji F, Wang L, et al. Asymptomatic and human-to-human transmission of SARS-CoV-2 in a 2-family cluster, Xuzhou, China. *Emerg Infect Dis*. 2020;26(7):1626-1628. https://doi.org/10.3201/eid2607.200718.
4. Bai Y, Yao L, Wei T, et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA*. 2020;323(14):1406. https://doi.org/10.1001/jama.2020.2565.
5. Pan A, Liu L, Wang C, et al. Association of Public Health Interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA*. 2020;323:1915. https://doi.org/10.1001/jama.2020.6130.
6. Railsback SF, Grimm V. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton, NJ: Princeton University Press; 2012.
7. Wilensky U, Rand W. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge, MA: MIT Press; 2015.
8. Eisinger D, Thulke H-H. Spatial pattern formation facilitates eradication of infectious diseases. *J Appl Ecol*. 2008;45(2):415-423. https://doi.org/10.1111/j.1365-2664.2007.01439.x.
9. Squazzoni F, Polhill JG, Edmonds B, et al. Computational models that matter during a global pandemic outbreak: a call to action. *J Artif Soc Soc Simul*. 2020;23(2):10. https://doi.org/10.18564/jasss.4298.
10. Sridhar D, Majumder MS. Modelling the pandemic. *BMJ*. 2020;369:m1567. https://doi.org/10.1136/BMJ.M1567.
11. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf.
12. OCEBM Levels of Evidence Working Group. *The Oxford 2009 Levels of Evidence*. Oxford: Oxford Centre for Evidence-Based Medicine;

2009 https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/.
13. National Institute for Health and Care Excellence. *Developing NICE Guidelines: the Manual*. London: National Institute for Health and Care Excellence; 2015 https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf.
14. Rocca E. The judgements that evidence-based medicine adopts. *J Eval Clin Pract*. 2018;24(5):1184-1190. https://doi.org/10.1111/jep.12994.
15. Worrall J. Evidence: philosophy of science meets medicine. *J Eval Clin Pract*. 2010;16(2):356-362. https://doi.org/10.1111/j.1365-2753.2010.01400.x.
16. Williamson J. EBM+: increasing the systematic use of mechanistic evidence. *Oral Presentations*. Vol 24. London: BMJ Publishing Group; 2019:A13-A14. https://doi.org/10.1136/bmjebm-2019-EBMLive.25.
17. Parkkinen V-P, Wallmann C, Wilde M, et al. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Berlin: Springer; 2018.
18. Clarke B, Gillies D, Illari P, Russo F, Williamson J. The evidence that evidence-based medicine omits. *Prev Med (Baltim)*. 2013;57(6):745-747. https://doi.org/10.1016/J.YPMED.2012.10.020.
19. Russo F, Williamson J. Interpreting causality in the health sciences. *Int Stud Philos Sci*. 2007;21(2):157-170. https://doi.org/10.1080/02698590701498084.
20. Marshall BDL, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. *Am J Epidemiol*. 2015;181(2):92-99. https://doi.org/10.1093/aje/kwu274.
21. Hernan MA. Invited commentary: agent-based models for causal inference—reweighting data and theory in epidemiology. *Am J Epidemiol*. 2015;181(2):103-105. https://doi.org/10.1093/aje/kwu272.
22. Clarke B, Gillies D, Illari P, Russo F, Williamson J. Mechanisms and the evidence hierarchy. *Topoi*. 2014;33(2):339-360. https://doi.org/10.1007/s11245-013-9220-9.
23. Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis*. 2020;12(3):165-174. https://doi.org/10.21037/jtd.2020.02.64.
24. Neher RA, Dyrdak R, Druelle V, Hodcroft EB, Albert J. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. *Swiss Med Wkly*. 2020;150(1112):w20224. https://doi.org/10.4414/smw.2020.20224.
25. Giordano G, Blanchini F, Bruno R, et al. A SIDARTHE Model of COVID-19 Epidemic in Italy. 2020. http://arxiv.org/abs/2003.09861. Accessed May 16, 2020
26. Acuna-Zegarra MA, Comas-Garcia A, Hernandez-Vargas E, Santana-Cibrian M, Velasco-Hernandez JX. The SARS-CoV-2 Epidemic Outbreak: a Review of Plausible Scenarios of Containment and Mitigation for Mexico. *medRxiv*. 2020;2020:20046276. https://doi.org/10.1101/2020.03.28.20046276.
27. Kissler SM, Tedijanto C, Lipsitch M, Grad Y. Social distancing strategies for curbing the COVID-19 epidemic. *medRxiv*. 2020;2020:20041079. https://doi.org/10.1101/2020.03.22.20041079.
28. Peng L, Yang W, Zhang D, et al. Epidemic Analysis of COVID-19 in China by Dynamical Modeling. 2020. http://arxiv.org/abs/2002.06563. Accessed May 16, 2020.
29. Fu X, Ying Q, Zeng T, Long T, Wang Y. Simulating and forecasting the cumulative confirmed cases of SARS-CoV-2 in China by Boltzmann function-based regression analyses. *J Infect*. 2020;80(5):578-606. https://doi.org/10.1016/J.JINF.2020.02.019.
30. Tobías A. Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Sci Total Environ*. 2020;725:138539. https://doi.org/10.1016/J.SCITOTENV.2020.138539.
31. Chang SL, Harding N, Zachreson C, et al. Modelling Transmission and Control of the COVID-19 Pandemic in Australia. 2020. http://arxiv.org/abs/2003.10218. Accessed May 16, 2020.
32. Adam D. Special report: the simulations driving the world's response to COVID-19. *Nature*. 2020;580(7803):316-318. https://doi.org/10.1038/d41586-020-01003-6.
33. Cliff OM, Harding N, Piraveenan M, Erten EY, Gambhir M, Prokopenko M. Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: an agent-based modelling approach. *Simul Model Pract Theory*. 2018;87:412-431. https://doi.org/10.1016/J.SIMPAT.2018.07.005.
34. Cauchemez S, Bhattarai A, Marchbanks TL, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc Natl Acad Sci*. 2011;108(7):2825-2830. https://doi.org/10.1073/PNAS.1008895108.
35. Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *J Med Virol*. 2020;92(6):675-679. https://doi.org/10.1002/jmv.25723.
36. Liu Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet*. 2020;395(10227):e47. https://doi.org/10.1016/S0140-6736(20)30462-1.
37. Illari PM, Williamson J. What is a mechanism? Thinking about mechanisms across the sciences. *Eur J Philos Sci*. 2012;2(1):119-135. https://doi.org/10.1007/s13194-011-0038-2.
38. Glennan S. *The New Mechanical Philosophy*. Oxford: Oxford University Press; 2017.
39. Glennan S, Illari PM. Varieties of mechanisms. In: Glennan S, Illari PM, eds. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. London: Routledge; 2018:91-103.
40. Kelly MP, Kelly RS, Russo F. The integration of social, behavioral, and biological mechanisms in models of pathogenesis. *Perspect Biol Med*. 2014;57(3):308-328. https://doi.org/10.1353/pbm.2014.0026.
41. Portides D. Idealization and abstraction in scientific modeling. *Synthese*. 2018;1-23. https://doi.org/10.1007/s11229-018-01919-7.
42. Frieden TR, Lee CT. Identifying and interrupting Superspreading events—implications for control of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis*. 2020;26(6):1059-1066. https://doi.org/10.3201/eid2606.200495.
43. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005;438(7066):355-359. https://doi.org/10.1038/nature04153.
44. Wong G, Liu W, Liu Y, Zhou B, Bi Y, Gao GF. MERS, SARS, and Ebola: the role of super-spreaders in infectious disease. *Cell Host Microbe*. 2015;18(4):398-401. https://doi.org/10.1016/J.CHOM.2015.09.013.
45. Everhart JE, Kruszon-Moran D, Perez-Perez GI, Tralka TS, McQuillan G. Seroprevalence and ethnic differences in helicobacter pylori infection among adults in the United States. *J Infect Dis*. 2000;181(4):1359-1363. https://doi.org/10.1086/315384.
46. Lazarus R, Vercelli D, Palmer LJ, et al. Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunol Rev*. 2002;190(1):9-25. https://doi.org/10.1034/j.1600-065X.2002.19002.x.
47. Delgado JC, Baena A, Thim S, Goldfeld AE. Ethnic-specific genetic associations with pulmonary tuberculosis. *J Infect Dis*. 2002;186(10):1463-1468. https://doi.org/10.1086/344891.
48. Redelings MD, Sorvillo F, Mascola L. Increase in *Clostridium difficile*-related mortality rates, United States, 1999–2004. *Emerg Infect Dis*. 2007;13(9):1417-1419. https://doi.org/10.3201/eid1309.061116.
49. Smith KGC, Clatworthy MR. FcγRIIB in autoimmunity and infection: evolutionary and therapeutic implications. *Nat Rev Immunol*. 2010;10(5):328-343. https://doi.org/10.1038/nri2762.
50. Zou L, Ruan F, Huang M, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med*. 2020;382(12):1177-1179. https://doi.org/10.1056/NEJMc2001737.

51. Andersen H. Mechanisms: what are they evidence for in evidence-based medicine? *J Eval Clin Pract*. 2012;18(5):992-999. https://doi.org/10.1111/j.1365-2753.2012.01906.x.
52. Bankes S. Exploratory modeling for policy analysis. *Oper Res*. 1993;41(3):435-449. https://doi.org/10.1287/opre.41.3.435.
53. Wu J, Dhingra R, Gambhir M, Remais JV. Sensitivity analysis of infectious disease models: methods, advances and their application. *J R Soc Interface*. 2013;10(86):20121018. https://doi.org/10.1098/rsif.2012.1018.
54. Weisberg M. Robustness analysis. *Philos Sci*. 2006;73(5):730-742. https://doi.org/10.1086/518628.
55. Machamer P, Darden L, Craver CF. Thinking about mechanisms. *Philos Sci*. 2000;67(1):1-25. https://doi.org/10.1086/392759.
56. Giere RN. How models are used to represent reality. *Philos Sci*. 2004;71(5):742-752. https://doi.org/10.1086/425063.
57. Giere RN. An agent-based conception of models and scientific representation. *Synthese*. 2010;172(2):269-281. https://doi.org/10.1007/s11229-009-9506-z.
58. Glennan S. Modeling mechanisms. *Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci*. 2005;36(2):443-464. https://doi.org/10.1016/J.SHPSC.2005.03.011.
59. Maziarz M. *The Philosophy of Causality in Economics: Causal Inferences and Policy Proposals*. London: Routledge; 2020.
60. Eubank S, Eckstrand I, Lewis B, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Bull Math Biol*. 2020;82(4):52. https://doi.org/10.1007/s11538-020-00726-x.
61. Worldometers.info. Coronavirus. https://www.worldometers.info/coronavirus/country/australia/. Accessed June 25, 2020.
62. Pannett R. Australia restricts travelers from mainland China as virus impact spreads. Wall Street Journal. https://www.wsj.com/articles/australia-s-qantas-suspends-china-flights-as-virus-impact-spreads-11580536238. 2020, February 1.
63. Worthington B, Snape J. Australia blocks arrival of all non-citizens, non-residents in expanded coronavirus travel ban. *ABC News*. https://www.abc.net.au/news/2020-03-19/coronavirus-non-resident-travel-ban-australia/12071640. 2020, March 19.
64. Australian Prime Minister. Update on coronavirus measures. https://www.pm.gov.au/media/update-coronavirus-measures-220320. 2020, March 22.
65. Knaus C, Wahlquist C, Remeikis A. PM Announces Pubs, Clubs and Cinemas to Close, Schools Stay Open in Stage One Measures—as it Happened. *The Guardian*. https://www.theguardian.com/world/live/2020/mar/22/coronavirus-updates-live-australia-nsw-victoria-qld-tasmania-cases-government-stimulus-latest-update-news. 2020, March 22.
66. National Cabinet. National Cabinet Statement. https://www.pm.gov.au/media/national-cabinet-statement. 2020, March 29.
67. Karp P, Davey M. Why Australia Is Not Shutting Schools to Help Control the Spread of Coronavirus. *The Guardian*. https://www.theguardian.com/world/2020/mar/16/why-australia-is-not-shutting-schools-to-help-control-the-spread-of-coronavirus. 2020, March 16.
68. Victoria State Government. Coronavirus update for Victoria. https://www.dhhs.vic.gov.au/coronavirus-update-victoria-22-june-2020. 2020, June 22.
69. Dignum F, Dignum V, Davidsson P, et al. Analysing the combined health, social and economic impacts of the corovanvirus pandemic using agent-based social simulation. 2020. http://arxiv.org/abs/2004.12809. Accessed May 16, 2020.
70. Bodenstein M, Corsetti G, Guerrieri L. Social distancing and supply disruptions in a pandemic. *Financ Econ Discuss Ser*. 2020;2020(031):1–52. https://doi.org/10.17016/feds.2020.031.
71. Broadbent A, Smart B. Why a one-size-fits-all approach to COVID-19 could have lethal consequences. 2020. https://theconversation.com/why-a-one-size-fits-all-approach-to-covid-19-could-have-lethal-consequences-134252.
72. Anjum RL, Copeland S, Rocca E. Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'. *BMJ Evidence-Based Med*. 2020;25(1):6-8. https://doi.org/10.1136/BMJEBM-2018-111092.
73. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine. How to Practice and Teach EBM*. 2nd ed. London: Churchill Livingstone; 2000.

How to cite this article: Maziarz M, Zach M. Agent-based modelling for SARS-CoV-2 epidemic prediction and intervention assessment: A methodological appraisal. *J Eval Clin Pract*. 2020;26:1352–1360. https://doi.org/10.1111/jep.13459