

# Complex systems models for causal inference in social epidemiology

Hiba N Kouser, Ruby Barnard-Mayers , Eleanor Murray 

► Supplemental material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2019-213052>).

Epidemiology, Boston University, Boston, Massachusetts, USA

## Correspondence to

Eleanor Murray, Department of Epidemiology, Boston University, Boston, MA 02445, USA; [ejmurray@bu.edu](mailto:ejmurray@bu.edu)

Received 24 June 2020

Revised 24 September 2020

Accepted 21 October 2020

## ABSTRACT

Systems models, which by design aim to capture multi-level complexity, are a natural choice of tool for bridging the divide between social epidemiology and causal inference. In this commentary, we discuss the potential uses of complex systems models for improving our understanding of quantitative causal effects in social epidemiology. To put systems models in context, we will describe how this approach could be used to optimise the distribution of COVID-19 response resources to minimise social inequalities during and after the pandemic.

## INTRODUCTION

The health and well-being of the public are determined by complex interactions between biological, individual, societal and environmental factors.<sup>1</sup> These often intertwine in dynamic and context-specific ways that can make studying any one component in isolation misleading or limiting. Feedback loops; leverage, tipping points or other non-linear relationships; interference, transmission, or spillover; and emergent properties such as social organisation can make traditional regression-based analyses of observational, or even randomised trial, data insufficient for understanding the causes and determinants of public health.<sup>2–9</sup> Social epidemiology has long grappled with this complexity and emphasises the need for formulating and formalising a clear theoretical framework before attempting to untangle these forces.<sup>1–10</sup> Agent-based models or compartmental models have been proposed as a solution for understanding complex systems in social epidemiology.<sup>4–6 11 12</sup>

In this paper, we review the use of these systems models in social epidemiology and link them to quantitative estimation of causal effects under the potential outcomes framework in the hopes of further advancing the use of complex systems methods in social epidemiology.

## Brief introduction to complex systems models

Complex systems are those with a large number of interacting components.<sup>5</sup> Complex systems models encompass a broad range of methodologies that generally use computational simulation of relationships between features in a pre-specified structure to understand the ways in which these features interact. In the context of public health, two major classes of models are of particular interest: group-level and unit-level systems models.

Group-level models include infectious disease transmission models such as SIR and SEIR models;

health policy models, such as Markov cohort models; and compartmental models of disease progression. The common feature of these models is that they simulate the evolution of the point prevalence of one or more conditions of interest within groups or strata of individuals. For example, in an SIR model, the strata of interest are ‘Susceptible’, ‘Infectious’, and ‘Recovered’, whereas in a policy model, strata of interest might include ‘Symptomatic’, ‘Diagnosed’, and ‘Connected to Care’. These strata are often referred to as ‘states’. Group-level models can be dynamic or static. In dynamic models, such as the SIR model, the number of people within a given state at a particular time point influences the rate of movement through the system of individuals in other states. In static models, such as a Markov cohort model, the rate of movement through the system depends only on an individual’s current state (possibly including an indicator for some key features such as age or gender), but not on the current states of other individuals. Group-level models have the advantage of being relatively low complexity and requiring minimal computational resources—indeed they can generally be solved numerically without simulation. However, the trade-off for computational simplicity is that these models can be challenging to design and parameterise. Limiting the number of states requires strong assumptions about which strata are meaningful to include and what factors govern the rate of progression between strata,<sup>13</sup> and determining appropriate values for these progression rates can be challenging especially when the population of interest has not been well-studied.

Unit-level models include agent-based models (often called ‘ABMs’) and individual-level simulation models (also called microsimulation models).<sup>14–16</sup> The defining feature of these models is that each simulated unit represents a single individual in a (virtual) population, as opposed to a group of individuals as modelled in the group-level models. The chief benefit of simulating individuals rather than groups is the ability to track and include aspects of an individual’s personal history, location in space and time relative to other individuals, and interactions between individuals. The main difference between agent-based models and individual-level simulation models is whether the individual units are simulated in isolation from each other (individual-level models), or whether units can transmit information (eg, education or infection) between one another (agent-based models); although this distinction is not always consistently defined in the literature.<sup>14</sup> This transmission allows for the possibility of interference, also called



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Kouser HN, Barnard-Mayers R, Murray E. *J Epidemiol Community Health* Epub ahead of print: [please include Day Month Year]. doi: 10.1136/jech-2019-213052

spillover or dependent happenings, which occurs when the outcomes of one or more individuals in the population affect the exposure and/or outcome of other individuals.<sup>17-19</sup> In addition to the interaction between agents, agent-based models can consider heterogeneity, or differences, among agents and types of agents within an environment. Heterogeneity is necessary to represent the complex populations that exist in the real-world and is incorporated by including in the agent-based model the types of actors that may exist in a system. Heterogeneity can be modelled at the exposure, outcome, covariate, or probability levels.<sup>19</sup>

Complex systems models are designed to represent key states, characteristics, or decisions that individuals or groups pass through over time. One common way of visualising the structure of a systems model is to create a decision-tree that depicts the relevant variables, the possible pathways, and the potential downstream consequences for an agent in the simulation. This structure can be used to depict both group-level and individual-level systems models, although group-level models are often more commonly depicted with compartment or flow diagrams. For simple systems models, the decision tree can also be used to run the simulation.<sup>20</sup>

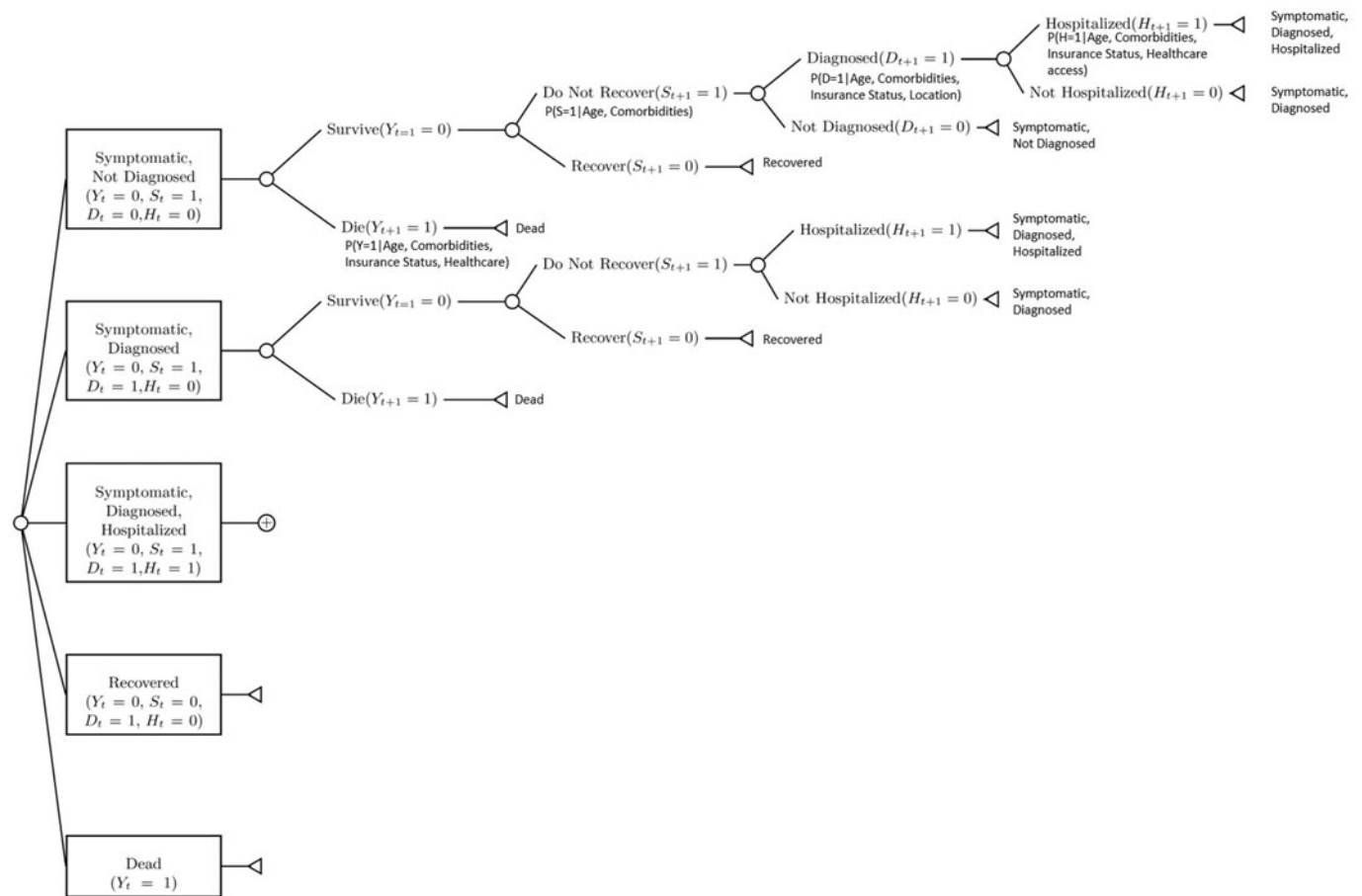
Figure 1 depicts a possible systems model for COVID-19 using the decision-tree framework. This tree describes five possible states at any given time for an individual with COVID-19: (1) currently symptomatic, but not yet diagnosed or hospitalised; (2)

currently symptomatic and diagnosed, but not yet hospitalised; (3) currently symptomatic, diagnosed and hospitalised; (4) recovered; and (5) dead. In order to evaluate the impact of social factors on the progression of individuals between these states, the decision tree allows the probability of transitioning between these states to depend on the individual's underlying health status, insurance status and access to testing.

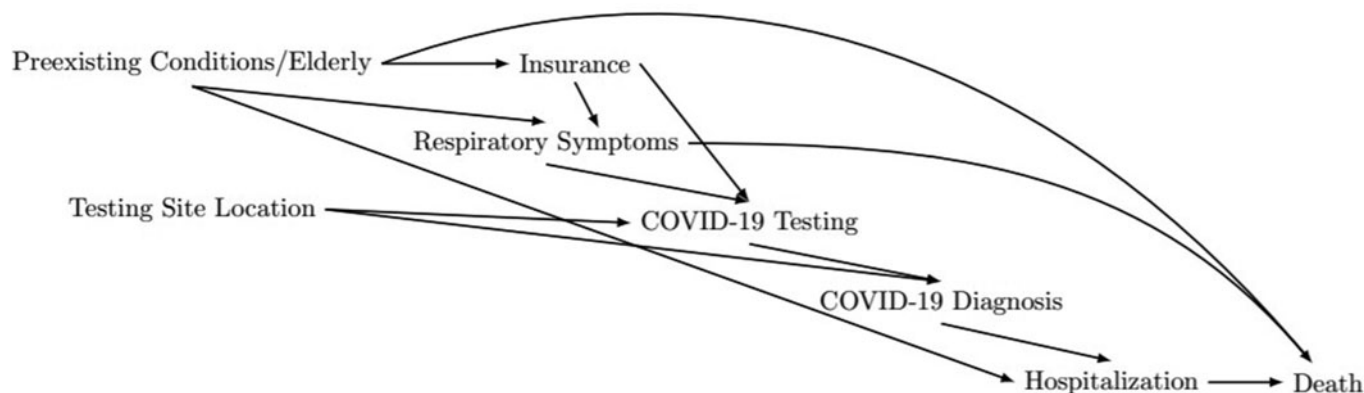
**Social epidemiologic example: inequity and COVID-19**

To put systems models in context, we use an example based on COVID-19 and related social inequalities that both exacerbate and are exacerbated by the 2020 pandemic (figure 2). Recent COVID-19 research has found large disparities in COVID-19 morbidity and mortality between minoritized and majority communities around the world.<sup>21-24</sup> The relationship between social epidemiology and infectious disease is a well-established one,<sup>1 25-27</sup> and many of the complexities encountered in social epidemiology have been studied in the infectious disease context. We hope to provide social epidemiologists with a toolkit to both understand the growing number of COVID-19 models and apply those modelling skills to other important social epidemiologic questions by using COVID-19 as an example.

A substantial body of literature exists highlighting the disproportionate impact of COVID-19 on People of Colour including Black and Indigenous communities, both internationally<sup>28 29</sup> and across



**Figure 1** Decision tree for a systems model evaluating the role of access to testing on COVID-19 morbidity and mortality. Testing access in the model is represented in the conditional probability of diagnosis given age, comorbidities, insurance status, and location. Letters represent variables and subscripts represent time-point and are included to allow the reader to gain intuition about how individuals move through the system: Y is vital status (0=alive, 1=dead); S is symptom status (0=no symptoms; 1= symptoms); H is hospital status (0=not hospitalised; 1=hospitalised); D is diagnostic status (0=not diagnosed, 1=diagnosed).



**Figure 2** Causal directed acyclic graph for the relationships between individual and neighbourhood characteristics, SARS-CoV-2 testing, and COVID-19 morbidity and mortality.

the United States.<sup>30–32</sup> In the US, Black and Indigenous communities have been particularly negatively affected by COVID-19 compared to non-Hispanic whites,<sup>30–40</sup> with exacerbated disparities at the intersection of age and race/ethnicity.<sup>34 41–45</sup>

Although some have argued that variation in underlying health status is to blame,<sup>46</sup> it is more likely the disparities in both health status and COVID-19 stem from wide-spread structural problems, including systemic and structural racism; racial capitalism, in which racial oppression is crucial for the accumulation of capital rooted in American culture,<sup>47</sup> and medical racism, in which minoritized individuals receive lower quality care and experience greater delays and barriers to accessing healthcare.<sup>48–60</sup>

For example, in the US racial and ethnic minorities make up a significant proportion of essential workers who cannot work from home<sup>57</sup> and often rely heavily on public transportation.<sup>58</sup> These factors put them at greater risk for acquiring the virus than those who can quarantine completely. Furthermore, 25% of the essential workers in the US live in low-income households, and 18% have at least one uninsured household member.<sup>59</sup> The differences in the allocation of public health infrastructure and resources create vulnerable communities that may not be well-supported to deal with the pandemic.<sup>60–65</sup>

For the remainder of this article, we explore how complex systems models could aid in answering the social epidemiologic question: ‘How can we best distribute limited COVID-19 testing resources in order to minimise social inequities?’

### DESIGNING SYSTEMS MODELS FOR CAUSAL INFERENCE

Systems models have long been used to generate hypotheses about possible causal relationships, triangulate information on potential mechanisms, and identify potentially important leverage points.<sup>66</sup> When our question of interest requires a quantitative estimate of a causal effect for decision making, systems models can also be useful but need to be defined to clearly and carefully specify the causal assumptions. The use of quantitative methods for estimating causal effects in social epidemiology has been the subject of much, sometimes heated, debate.<sup>10 67–80</sup> However, many of these debates revolve around the problem of specifying sufficiently complex causal questions to capture many levels at which social determinants of health operate. We argue that systems models, which by design aim to capture multi-level complexity, are in fact the natural choice of tool for bridging the divide between these two sub-disciplines, especially as the use of these models for estimating counterfactual outcomes is becoming more formalised.<sup>6 9 81–84</sup>

The key feature of designing systems models for quantitative causal effect estimation is to design the model with a clear set of causal assumptions. The required assumptions are similar to those required for any causal effect estimation using randomised trials or observational data,<sup>81</sup> with the additional caveat that we must now be sure to *design and parameterise* our model correctly, not just measure and adjust appropriately. We briefly review the required causal assumptions for traditional methods and describe special considerations for systems models.

The first assumption is the Stable Unit Treatment Value Assumption, or SUTVA. This assumption can roughly be decomposed into two components: consistency and no interference.

Consistency requires that the counterfactual world we think we are making inference about is the one we are actually making inference about. In practice, this is typically supported by the need for a clear causal question, reflecting a comparison between two or more interventions or exposure states applied to a specific group of individuals,<sup>85</sup> although the appropriate formulation of these questions in social epidemiology has been questioned.<sup>86</sup> In the context of a complex system, observational data and/or randomised trials are often unlikely to be sufficient to establish consistency since the data requirements for unravelling complexity increase rapidly. In contrast, complex systems models can allow us to be more creative in our specification of causal contrasts, more easily apply multi-level interventions, and assess multi-level outcomes. Creating and using a systems model requires clearly specifying a framework under which key characteristics are related, and an algorithm by which the desired contrasting exposure levels can be obtained. Although these may not always clearly represent real-world policies, the algorithm must be detailed enough for the computer (and by extension the researcher) to thoroughly understand the counterfactual scenarios of interest. This detailed algorithm provides a specification for the meaning of the counterfactuals we aim to compare—that is, it allows us to assume that our results are *at minimum consistent* for the exposure contrast we have specified in the simulated world which the model represents.

For example, if we are interested in understanding whether a COVID testing strategy based on prioritising individuals in high-incidence neighbourhoods helps to reduce racial and ethnic disparities in COVID-19 hospitalisation, using observational or trial data may be difficult due to the complex interplay between structural racism, COVID risk and vulnerability, residential neighbourhood, and healthcare access. In a complex systems model, we explicitly describe how we

believe these relationships work, and then assess the impact of our testing strategy *conditional* on our assumptions about the model structure and rules.

SUTVA also relies on the assumption of no interference—that is, that the exposures and outcomes of two or more individuals are independent of one another. By nature, many of the questions of interest that are assessed using systems models explicitly violate this assumption. However, there is a growing body of work that describes how to estimate causal effects in the presence of interference,<sup>17 18 87 88</sup> and these can be extended to complex systems models.<sup>89</sup>

The second key assumption of causal inference is positivity; all individuals in our population must have some probability of being in either group—no individual can be a ‘never treat’ or ‘always treat’ without a clear rule specifying this behaviour in the model. This can be difficult to achieve in observational data, but is built into the creation of systems models—all individuals must have a pre-specified probability of receiving an exposure, and researchers can assess the model to ensure there are no groups for which exposure will always or never be assigned. The requirement for positivity is a challenge for the use of the potential outcomes framework in social epidemiology—under this framework we can only estimate causal effects when data exist and exposures are possible. However, the use of complex systems models can allow us to somewhat relax this assumption by changing the rules of our simulated model universe to change the probability of exposure in a group for which exposure is actually always or never available. Note however that this requires strong assumptions about the impact of exposure in this group that may be difficult or impossible to justify.

The third key assumption is that of exchangeability—that is, the counterfactual distributions (and thus the distribution of covariates that cause confounding) for individuals who are in the exposed or treated group must be the same as those in the reference category.<sup>90 91</sup> These assumptions require being clear about the population of interest, and understanding all shared causes of our exposure and outcome of interest. In observational studies, exchangeability is supported via the collection of rich covariate data to ensure all confounding paths between exposure and outcome can be blocked.<sup>90</sup> The same principle can be applied to systems models with a slight modification: we must model all covariates necessary to ensure that all confounding paths between *all pairs of variables already in the systems model* can be blocked.<sup>81</sup> In addition, systems models require the additional assumption that the parameter inputs used to build the model are estimates of causal effects and were obtained from populations where these effects have the same value as in the population of interest.<sup>81 82</sup> This last assumption amounts to requiring the causal inference assumptions for every pair of variables in the systems model, and is often the most challenging to establish.<sup>82</sup>

One useful way of clarifying the above assumptions when designing a systems model is to employ the Target Trial Framework.<sup>92–94</sup> This framework asks the researcher to detail the protocol of an ideal (and sometimes unrealistic) randomised trial for the research question of interest. This helps define the specific features of the population, the specific comparison of two or more intervention or exposure groups, and the outcome(s) measurements of interest. This protocol then serves as the base on which to build a model, and allows other researchers to more easily assess whether they believe

the assumptions required are sufficient to interpret the results causally.

### Target trials in the presence of interference

An important consideration for systems models is that the ideal trial for many of the questions of interest would be one that asks complex questions such as ‘*how do we best distribute limited testing supplies to minimise social inequalities?*’. Questions such as this involve understanding that resources, information and exposure are shared between individuals, but their complexity does not prevent us from evaluating them using systems models, nor does it prevent us from specifying clearly defined, although complex, causal questions. For example, in the setting of SARS-CoV-2 testing, when a test is used on one individual it cannot be used on another individual. This problem can lead to challenges with the consistency assumption *unless* the causal question is specified in a way that accounts for the concepts of interference, transmission, spillover, or dependent happenings.

Fortunately, there are many existing solutions for how to ask complex causal questions under these settings. For example, complex trial designs have been developed which can allow researchers to answer these questions, including cluster-randomised, network-randomised, and ring vaccination trials.<sup>17 87 95</sup> In addition, the recently developed auto-g-computation algorithm could be extended to improving causal effect estimation with ABMs when the no interference assumption is violated.<sup>96</sup> Systems models can be designed to incorporate transmission of information (disease or resources) between individuals, and emulate the results of these more complex randomised trials to obtain an estimate of the causal effect under transmission scenarios.<sup>84</sup> These models are commonly used in infectious disease epidemiology where researchers can model outbreaks and the spread of disease due to the interaction between agents. This is especially relevant for COVID-19 models, and numerous research groups have already released models describing the extent and duration of the pandemic based on varying transmission assumptions.<sup>97 98</sup> But agent-based models can also be a valuable tool for understanding social epidemiologic questions. There are many features of the social environment, which transmit between individuals (eg, education, or violence), and many ways in which the outcomes or exposures of one individual can affect the exposure or outcome of another individual.

As an example, let’s revisit our decision-tree model depicted in figure 1. There, we allowed the availability of tests to influence the probability that an individual with COVID19 symptoms would be diagnosed. If our population of interest has access to only a fixed amount of tests over time, then we can incorporate this into our model by allowing the probability of testing to decrease over time, or by explicitly modelling the availability of tests—decreasing the number of tests as they are used, and increasing them when new supply is available. This second approach would require building an additional layer into our model that simulates the supply of tests over time. If relevant, a third layer could be added with information on location of tests or individuals to allow test access to vary over space as well as time.

We provide a step-by-step guide on how to build a systems model to answer our research question ‘How can we best distribute limited COVID-19 testing resources in order to minimise social inequities?’ in the online appendix. The first step to creating a systems model is to define a clear research question of interest, including a well-specified population, exposure and

reference groups, time period and location information, and a well-defined outcome of interest. The next step is to use the target trial framework to design a hypothetical trial protocol to answer the research question. Step 3 entails specifying variables and assumptions that will govern how agents interact and move through time/space, including potential confounders, colliders, mediators or effect modifiers. After this information has been agreed upon, the next step (Step 4) is to build the model to answer the research questions (from Step 1) using the trial protocol created in Step 2. Finally, Step 5 is to design and run additional analyses to test the accuracy and estimate the uncertainty of the model online appendix. At all steps when building a systems model, it is important to keep in mind the parameters and rules to ensure a good fit.

## CONCLUSION

Systems models are a valuable tool for understanding the complexity of public health as it relates to societal factors and distribution of disease.<sup>6 15 16 19</sup> These models provide us with a tool for assessing relationships between multi-level features, including both individual, community and societal factors. Importantly, other methods, such as DAG-based regression modelling, can and should also tackle these issues when data exist or can be directly collected. The chief benefit of systems models is that they can be used in the absence of a complete or rich dataset on the specific target population of interest. Recently, methodological developments have increased the ease with which these models can be designed and implemented when estimation of causal effects is of interest. With this commentary, we hope to provide readers with an overview of how these recent developments work and can be applied to sociological questions.

We presented an example of a model assessing a social epidemiological question about COVID-19 mortality: *how do we best distribute SARS-CoV-2 test access in order to minimise social inequities*. This question could be further refined to require that we not only identify the lowest disparities based on race and ethnicity but also by other demographic variables—including age, gender, disability status or employment type. The model we discussed focussed only on identifying and treating those who were already ill with COVID-19, but it could also be extended to consider the transmission of SARS-CoV-2 between individuals, or account for geographical variations in transmission, access to care, or access to testing resources.

Systems models are being increasingly used for social and chronic disease epidemiology, and their use entails many benefits to the researcher. Unlike observational data analyses, systems models can be increased in complexity with relatively little additional data requirements. However, these lower data requirements come at the cost of requiring more assumptions in order to imbue the resulting numeric answer with a meaningful causal interpretation. It is important that researchers are clear and explicit about the assumptions they are making when designing and using these models. The target trial framework, coupled with the use of causal directed acyclic graphs, provides a relatively simple toolkit for specifying these assumptions and providing reproducible and reliable answers from systems models.

**Twitter** Eleanor Murray @epiellie, Hiba Kouser @h\_kouser and Ruby Barnard-Mayers @rubymaei.

**Acknowledgements** This project was supported by the NICHD (R21HD098733).

**Contributors** All authors contributed to the conceptualization and drafting of the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Commissioned; externally peer reviewed.

**Data availability statement** Data sharing not applicable as no datasets generated and/or analysed for this study.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the COVID-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

## ORCID iDs

Ruby Barnard-Mayers <http://orcid.org/0000-0003-2870-6614>

Eleanor Murray <http://orcid.org/0000-0003-0043-4901>

## REFERENCES

- Krieger N. *Epidemiology and the people's health*. Oxford: Oxford University Press, 2011.
- Olaya C. Cows, agency, and the significance of operational thinking. *Syst Dyn Rev* 2015;31:183–219.
- Hovmand PS. Community based system dynamics [Internet]. New York, NY: Springer New York, 2014. Available <http://link.springer.com/10.1007/978-1-4614-8763-0> (accessed 25 Aug 2020)
- Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. *Am J Epidemiol* 2008;168:1–8.
- Diez Roux AV. Integrating social and biologic factors in health research: a systems view. *Ann Epidemiol* 2007;17:569–74.
- Marshall BDL, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. *Am J Epidemiol* 2015;181:92–9.
- Bradley E. Causes and effects of chaos. *Report No.: A.I. Memo No. 1216*. Massachusetts Institute of Technology Artificial Intelligence Laboratory, December 1990: 45.
- Leverage points: places to intervene in a system [Internet]. *The Academy for Systems Change*. Available <http://donellameadows.org/archives/leverage-points-places-to-intervene-in-a-system/> (accessed 25 Aug 2020)
- Hernán MA. Invited commentary: agent-based models for causal inference: reweighting data and theory in epidemiology. *Am J Epidemiol* 2015;181:103–5.
- Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol* 2016;45:1787–808.
- de Marchi S, Page SE. Agent-based Models. *Annu Rev Polit Sci*. 2014 May;17:1–20
- Tracy M, Cerdá M, Keyes KM. Agent-based modeling in public health: current applications and future directions. *Annu Rev Public Health* 2018;39:77–94.
- Huinink MGM. *Decision making in health and medicine: integrating evidence and values*. Cambridge; New York: Cambridge University Press, 2001: 305–38. Available <http://hollis.harvard.edu/?itemid=%7Clibrary/m/aleph%7C008748782>
- Arnold KF, Harrison WJ, Heppenstall AJ, et al. DAG-informed regression modelling, agent-based modelling and microsimulation modelling: a critical comparison of methods for causal inference. *Int J Epidemiol* 2019;48:243–53.
- Bruch E, Atwell J. Agent-based models in empirical social research. *Social Methods Res* [Internet]. 24 October 2013. Available <https://journals.sagepub.com/doi/10.1177/0049124113506405> (accessed 13 May 2020)
- Badham J, Chattoe-Brown E, Gilbert N, et al. Developing agent-based models of complex health behaviour. *Health Place* 2018;1:170–7.
- Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiol Camb Mass* 1991;2:331–8.
- Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology* 1995;6:142–51.
- El-Sayed AM, Scarborough P, Seemann L, et al. Social network analysis and agent-based modeling in social epidemiology. *Epidemiol Perspect Innov* 2012;9:1.
- TreeAge Pro 2020, R2. *TreeAge Software*.
- Pareek M, Bangash MN, Pareek N, et al. Ethnicity and COVID-19: an urgent public health research priority. *Lancet* 2020;395:1421–2.
- Hooper MW, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA* 2020 Jun 23 323; 2466–7.

- 23 Brandt EB, Beck AF, Mersha TB. Air pollution, racial disparities, and COVID-19 mortality. *J Allergy Clin Immunol* [Internet]. 2020 May 7;146:61–3; (accessed 23 Jun 2020);
- 24 Laurencin CT, The MA. COVID-19 pandemic: a call to action to identify and address racial and ethnic disparities. *J Racial Ethn Health Disparities* 2020;7:398–402.
- 25 Snow J. *On the mode of communication of cholera*. John Churchill, 1855: 216.
- 26 Nowotny K, Bailey Z, Omori M, et al. Exposures need for progressive criminal justice reform. *Am J Public Health* 2020;110:967–8.
- 27 Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. *Am J Epidemiol* 1999;150:113–20.
- 28 Bhalal N, Curry G, Martineau AR, et al. Sharpening the global focus on ethnicity and race in the time of COVID-19. *Lancet Lond Engl* 2020 30; 395: 1673–6.
- 29 Martinez-Juarez LA, Sedas AC, Orcutt M, et al. Governments and international institutions should urgently attend to the unjust disparities that COVID-19 is exposing and causing. *EClinicalMedicine* 2020;23:100376.
- 30 Yancy CW. COVID-19 and African Americans. *JAMA* 2020 Apr;15.
- 31 Cyrus E, Clarke R, Hadley D, et al. The impact of COVID-19 on African American communities in the United States. *MedRxiv Prepr Serv Health Sci* 2020 May, 19.
- 32 Rentsch CT, Kidwai-Khan F, Tate JP, et al. COVID-19 by race and ethnicity: a national cohort study of 6 million United States veterans. *MedRxiv Prepr Serv Health Sci* 2020 May 18.
- 33 Rosenberg ES, Tesoriero JM, Rosenthal EM, et al. Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. *Ann Epidemiol* 2020 Jun 16.
- 34 Bassett M, Chen JT, Krieger N. The unequal toll of COVID-19 mortality by age in the United States: quantifying racial/ethnic disparities. *Harv Cent Popul Dev Ser* 2020;19:3.
- 35 Price-Haywood EG, Burton J, Fort D, et al. Hospitalization and mortality among black patients and white patients with COVID-19. *N Engl J Med* 2020 25; 382: 2534–43.
- 36 Sehra ST, Fundin S, Lavery C, et al. Differences in race and other state-level characteristics and associations with mortality from COVID-19 infection. *J Med Virol* 2020 May;30.
- 37 Martinez DA, Hinson JS, Klein EY, et al. SARS-CoV-2 positivity rate for Latinos in the Baltimore-Washington, DC Region. *JAMA* 2020 Jun 18;
- 38 Kakol M, Upson D, Sood A. Susceptibility of Southwestern American Indian tribes to coronavirus disease 2019 (COVID-19). *J Rural Health* 2020 Apr 18.
- 39 Kaholokula JK, Samoa RA, Miyamoto RES, Palafox N, Daniels S-A. COVID-19 special column: COVID-19 hits native Hawaiian and Pacific Islander communities the hardest. *Hawaii J Health Soc Welf* 2020 1;79:144–6. PMID: PMC7226312.
- 40 Hu J-R, Wang M, Lu F. COVID-19 and Asian American Pacific Islanders. *J Gen Intern Med* Jun 2020;12.
- 41 CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) - United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Mar 27 69; 343–6.
- 42 Roxby AC, Greninger AL, Hatfield KM, et al. Outbreak investigation of COVID-19 among residents and staff of an independent and assisted living community for older adults in Seattle, Washington. *JAMA Intern Med* 2020 May 21;
- 43 Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N Engl J Med* 2020;28 382:2081–90.
- 44 Li Y, Temkin-Greener H, Gao S, et al. Infections and deaths among Connecticut nursing home residents: facility correlates. *J Am Geriatr Soc* 2020 Jun 18;
- 45 Ebor MT, Loeb TB, Trejo L. Social workers must address intersecting vulnerabilities among noninstitutionalized, black, latinx, and older adults of color during the COVID-19 pandemic. *J Gerontol Soc Work* 2020 Jun; 19;1–4.
- 46 Raifman MA, Raifman JR. Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income. *Am J Prev Med* 2020;59:137–9.
- 47 Laster Pirtle WN. Racial capitalism: a fundamental cause of novel coronavirus (COVID-19) pandemic inequities in the United States. *Health Educ Behav* 2020;47:504–8.
- 48 Bailey ZD, Krieger N, Agénor M, et al. Structural racism and health inequities in the USA: evidence and interventions. *Lancet Lond Engl* 2017 08;389:1453–63.
- 49 Bailey ZD, Moon JR. Racism and the political economy of COVID-19: will we continue to resurrect the past? *J Health Polit Policy Law* 2020 May 28;
- 50 Frederickson G. *Racism: a short history*. Princeton, NJ: Princeton University Press, 2003.
- 51 Zinn H. *A people's history of the United States: 1492: present*. New York, NY: Harper Perennial, 2015.
- 52 Dunbar-Ortiz R. *An indigenous peoples' history of the United States*. Boston, MA: Beacon Press, 2015.
- 53 Rollston R, Galea S. COVID-19 and the social determinants of health. *Am J Health Promot AJHP* 2020;34:687–9.
- 54 Krishnan L, Ogunwole SM, Cooper LA. Historical insights on coronavirus disease 2019 (COVID-19), the 1918 influenza pandemic, and racial disparities: illuminating a path forward. *Ann Intern Med* 2020 June 5;
- 55 Zhang CH, Schwartz GG. Spatial disparities in coronavirus incidence and mortality in the United States: an ecological analysis as of May 2020. *J Rural Health* 2020;36:433–45.
- 56 Wadhwa RK, Wadhwa P, Gaba P, et al. Variation in COVID-19 hospitalizations and deaths across New York City boroughs. *JAMA* 2020 Apr 29.
- 57 McNicholas C, Poydock M. Who are essential workers? A comprehensive look at their wages, demographics, and unionization rates [Internet]. *Economic Policy Institute: Working Economic Blog*. 2020. Available <https://www.epi.org/blog/who-are-essential-workers-a-comprehensive-look-at-their-wages-demographics-and-unionization-rates/> (accessed 11 Jul 2020)
- 58 Anderson M Who relies on public transit in the U.S. [Internet]. Pew Research Center, 2016. Available <https://www.pewresearch.org/fact-tank/2016/04/07/who-relies-on-public-transit-in-the-u-s/> (accessed 11 Jul 2020)
- 59 McCormack G, Avery C, Spitzer AK-L, et al. Economic vulnerability of households with essential workers. *JAMA* Jun, 2020 18.
- 60 Waltenburg MA, Victoroff T, Rose CE, et al. Update: COVID-19 among workers in meat and poultry processing facilities - United States, April–May 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:887–92.
- 61 Ramachandran P, Swamy L, Kaul V, et al. Strategy for ventilator and ICU resource allocation during the COVID-19 pandemic. *Chest* 2020May 12.
- 62 Sharma S, Cain J, Sakhuja A, et al. Guidance for healthcare providers managing COVID-19 in rural and underserved areas. *J Racial Ethn Health Disparities* 2020 Jun 10.
- 63 The unique H-SC, impact of COVID-19 on older adults in rural areas. *J Aging Soc Policy* 2020;32:396–402.
- 64 Essien UR, Eneanya ND, Crews DC. Prioritizing equity in a time of scarcity: the COVID-19 Pandemic. *J Gen Intern Med* 2020 June 30.
- 65 Farrell TW, Francis L, Brown T, et al. Rationing limited healthcare resources in the COVID-19 era and beyond: ethical considerations regarding older adults. *J Am Geriatr Soc* 2020;68:1143–9.
- 66 Galea S, Hall C, Kaplan GA. Social epidemiology and complex system dynamic modelling as applied to health behaviour and drug use research. *Int J Drug Policy* 2009;20:209–16.
- 67 Galea S, Hernán MA. Win-win: reconciling social epidemiology and causal inference. *Am J Epidemiol* 2020;189:167–70.
- 68 Schisterman EF. Editorial: let's be causally social. *Am J Epidemiol* 2020;189:165–6.
- 69 Jackson JW, Arah OA, Invited commentary: making causal inference more social and (social) epidemiology more causal. *Am J Epidemiol* 2020;189:179–82.
- 70 Robinson WR, Bailey ZD. Invited commentary: what social epidemiology brings to the table: reconciling social epidemiology and causal inference. *Am J Epidemiol* 2020;189:171–4.
- 71 Galea S, Hernán MA. Galea and Hernán respond to 'brings to the table,' 'differential measurement error,' and 'causal inference in social epidemiology'. *Am J Epidemiol* 2020;189:183–4.
- 72 VanderWeele TJ. Invited commentary: counterfactuals in social epidemiology: thinking outside of 'the box'. *Am J Epidemiol* 2020;189:175–8.
- 73 Daniel RM, De Stavola BL, Vansteelandt S. Commentary: the formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *Int J Epidemiol* 2016;45:1817–29.
- 74 VanderWeele TJ. Commentary: on causes, causal inference, and potential outcomes. *Int J Epidemiol* 2016;45:1809–16.
- 75 Blakely T, Lynch J. Commentary: BR DAGs and the restricted potential outcomes approach are tools, not theories of causation. *Int J Epidemiol* 2016;45:1835–7.
- 76 Robins JM, Weissman MB. Commentary: counterfactual causation and streetlamps: what is to be done? *Int J Epidemiol* 2016;45:1830–5.
- 77 Weed DL. Commentary: causal inference in epidemiology: potential outcomes, pluralism and peer review. *Int J Epidemiol* 2016;45:1838–40.
- 78 Naimi AI. On wagging tales about causal inference. *Int J Epidemiol* 2017 1;46:1340–2.
- 79 Krieger N, Smith GD. Reply to Naimi. *Int J Epidemiol* 2017 1;46:1342.
- 80 Krieger N, Davey Smith G. Response: FACEing reality: productive tensions between our epidemiological questions, methods and mission. *Int J Epidemiol* 2016;45:1852–65.
- 81 Murray EJ, Robins JM, Seage GR, et al. A comparison of agent-based models and the parametric g-formula for causal inference. *Am J Epidemiol* 2017;186:131–42.
- 82 Murray EJ, Robins JM, Seage GR, et al. The challenges of parameterizing direct effects in individual-level simulation models. *Med Decis Mak Int J Soc Med Decis Mak* 2020;40:106–11.
- 83 Murray EJ, Robins JM, Seage GR, et al. Using observational data to calibrate simulation models. *Med Decis Making* 2018;38:212–24.
- 84 Buchanan A, King M, Bessey S, et al. Disseminated effects in agent based models: a potential outcomes framework to inform pre-exposure prophylaxis coverage levels for HIV prevention. Society for Epidemiologic Research Annual Conference. Minneapolis, MN, 2019.
- 85 Hernan MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol* 2016;26:674–80.
- 86 Schwartz S, Prins SJ, Campbell UB, et al. Is the 'well-defined intervention assumption' politically conservative? *Soc Sci Med* 1982 2016;166:254–7.
- 87 Hudgens MG, Halloran ME, Toward causal inference with interference. *J Am Stat Assoc* 2008;103:832–42.
- 88 Buchanan AL, Vermund SH, Friedman SR, et al. Assessing individual and disseminated effects in network-randomized studies. *Am J Epidemiol* 2018;187:2449–59.
- 89 Buchanan A, Bessey S, Goedel WC, et al. Disseminated effects in agent based models: a potential outcomes framework to inform pre-exposure prophylaxis coverage levels for HIV prevention. *Am J Epidemiol* 2020.
- 90 Hernan MA, Robins J. *Causal inference: what if*. Boca Raton: Chapman & Hill/CRC, 2020.

- 91 Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol* 2010;171:674–7.
- 92 Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond. *Eur J Epidemiol* 2017;32:473–5.
- 93 García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol* 2017;32:495–500.
- 94 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016 15;183:758–64.
- 95 Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res* 2012;21:55–75.
- 96 Tchetgen EJT, Fulcher IR, Shpitser I. Auto-G-computation of causal effects on a network. *J Am Stat Assoc* 2020;0:1–33.
- 97 Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 2020 June;8:1–8.
- 98 SMDM COVID-19 modeling portal. *Smartsheet.com* [Internet]. Available <https://app.smartsheet.com/b/publish?EQBCT=1a3bc6acad99475f99acfd55a04a1564> (accessed 2020 Jun 23)