

RESEARCH

Situating Agent-Based Modelling in Population Health Research

Eric Silverman^{1*}, Umberto Gostoli¹, Stefano Picascia¹, Jonatan Almagor¹, Mark McCann¹, Richard Shaw¹ and Claudio Angione²

*Correspondence:

eric.silverman@glasgow.ac.uk

¹MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, 99 Berkeley Street, G3 7HR, Glasgow, UK

Full list of author information is available at the end of the article

Abstract

Today's most troublesome population health challenges are often driven by social and environmental determinants, which are difficult to model using traditional epidemiological methods. We agree with those who have argued for the wider adoption of agent-based modelling (ABM) in taking on these challenges. However, while ABM has been used occasionally in population health, we argue that for ABM to be most effective in the field it should be used as a means for answering questions normally inaccessible to the traditional epidemiological toolkit. In an effort to clearly illustrate the utility of ABM for population health research, and to clear up persistent misunderstandings regarding the method's conceptual underpinnings, we offer a detailed presentation of the core concepts of complex systems theory, and summarise why simulations are essential to the study of complex systems. We then examine the current state of the art in ABM for population health, and propose they are well-suited for the study of the 'wicked' problems in population health, and could make significant contributions to theory and intervention development in these areas.

Keywords: Agent-Based Modelling; Population Health; Complexity

1 Introduction

Health policy-making is, at the root, an attempt to undertake principled decisions in an environment of high uncertainty and high risk, in which competing pressures and interests from stakeholders have a significant effect on the actions ultimately taken [1]. The rise of causal inference methods (CIM), has attempted to address this by providing a framework under which we may predict the outcomes of proposed interventions.

However, CIM as applied in epidemiology today, has coalesced around a set of tools with certain limitations when applied to complex systems. For example, directed acyclic graphs (DAGs) are frequently used in CIM, but DAGs are unsuitable for modelling systems containing feedback loops (given they are acyclic), a common feature of complex social systems. Critics of CIM also suggest it takes an overly linear view of the decision-making process leading from evidence to decision-making, and have called for alternative concepts of cause that are not solely based on probabilistic statements about population outcomes in alternative worlds [2]. Decisions over causality can be based on pragmatic pluralism [3], or inference to the best explanation [4, 2] (often characterised as a form of abductive reasoning). Decision-makers ultimately must make decisions, even while accepting that their evidence

is incomplete or flawed or both, and that the exact causal process underlying the system of interest is still uncertain.

Here we propose that the addition of Agent-Based Modelling (ABM) and related complex-systems-based approaches to the population health toolkit will enable better-informed and more robust decision-making in population health. ABMs allow for the representation of causal processes in systems that include feedback loops and multiple layers of complex, interacting components. ABMs can model explicitly the individual-level decision-making that can lead to unexpected *emergent* effects at the population level. The necessity of constructing informed conceptual models of system processes when building an ABM also opens up new avenues for co-production of models with decision-makers, while also facilitating the inclusion of both quantitative and qualitative data. Taken in combination, these properties of ABMs enable them to serve as a complementary approach to the current population health toolset, and in turn to enable both scientists and policy makers to better navigate the non-linear abductive process leading from scientific evidence to policy action.

2 A new paradigm for complex public health

While ABMs offer great promise for the population health community, we suggest that their utility is too often evaluated from the viewpoint of mainstream traditional epidemiological methods [5, 6]. In addition, some population health models described as ‘agent-based’ would be better categorised as ‘individual-based’, as individuals are represented but display little or no capacity for agency or interaction [7, 8]. These issues have resulted in ABM studies being isolated experiments within the discipline, as their strengths are not being properly utilised, rather than developing into a mature category of population health research in their own right.

However, ABMs can have important and distinct uses in public health, exploring research questions that traditional methods cannot usefully address. ABMs allow us to understand how the combination of agent interactions and non-linear causal effects give rise to system-level patterns. Such an approach falls outside the traditional epidemiological toolkit, but developing our understanding of these patterns is essential for improving the health of the public. With this in mind we would concur with Auchincloss and Diez-Roux, who outlined how regression approaches can limit our enquiries:

Too often, the exclusive use of regression approaches constrains not only the answers we get but also the types of questions we pose and the hypotheses and even theoretical explanations that we develop. In our search for what is ‘tractable’ in empirical observational research (essentially that which mimics the perfect experiment), our questions have the tendency to become narrower and narrower and perhaps less relevant to understanding or intervening in the real world [9, p. 6].

Long after this paper was first published, the methodological debate in population health continues to discuss the relative merits of ABMs compared to regression in the presence of interference or dynamic processes [10]. In our view, the debate should acknowledge that the primary advantage of a model-centred approach is that it

allows us to answer fundamentally *different* types of questions. Consider the ability of ABMs to integrate variables of disparate nature, including qualitative, narrative, and anecdotal [11]; to enable the inclusion of geographic and social space at multiple levels of resolution – embedding feature-rich GIS representations, for instance, or sophisticated social network structures; and to simulate multiple complex processes simultaneously, all while remaining tractable (to a certain extent).

These features enable the exploration of processes in a fundamentally different way, and the examination of fundamentally different questions of wider scope and breadth. Simply put, adopting ABMs requires not just a change of methods, but a change of mindset and conceptual framework.

3 The need for a new approach to decision-making

Research in population health has to its credit a number of major successes over the years, due to concerted action on communicable diseases through improved sanitation and mass immunisation programmes, prompting major progress on serious health problems such as lung cancer and heart disease and a steady increase in life expectancy until recently [12]. More recently, significant progress has been made on noncommunicable diseases through tobacco control and lifestyle education. However, ‘despite major investment in both research and policy, effective action to tackle pressing contemporary public health challenges remains elusive’ [13]. The most intractable health problems are those that spring fundamentally from more complex causes: behavioural and social influence; and environmental interaction. Some notable ‘wicked’ problems [14] include obesity, alcohol and drug misuse, and the persistence of health inequalities, among others.

The ‘wicked’ health problems of the 21st century are driven by numerous influences where the causal links are not clearly defined, and the mechanisms that influence them are elusive [15]. Wicked problems in health are serious challenges for policy makers, given that such problems ‘are continually evolving; have many causal levels; have no single solution that applies in all circumstances and solutions can only be classified as better or worse, rather than right or wrong’ [16]. Behavioural risk factors underpin these wicked health problems, but the evidence is sparse as to how physical and social environments influence health behaviours, and what can be done to improve them.

Identifying, implementing and evaluating effective responses to major population health challenges requires a wider set of approaches beyond the traditional methods of public health research and should involve a wider set of actors beyond the health services [17]. Population health is influenced by multiple interacting determinants, including social, political, environmental, biological and behavioural factors. Current challenges to the health of the public, and the overarching inequalities in health within and across populations, are resistant to simple, linear, silver-bullet approaches [13].

Traditional epidemiological methods face the most difficulties in these types of population health challenges, in which determinants of illness and disease are multifaceted, inter-related and non-linear [18]. Unravelling the complex interactions of social and environmental determinants is challenging when using statistical methods

in which individuals and their actions may not be explicitly represented. Individual-based modelling practices such as microsimulation can model some of these behavioural and spatial effects, but the effects of interactions between individuals and environmental factors are still hard to capture.

We claim that the study of ‘wicked’ health problems necessitates the investigation of human society as a complex system, and that public health itself can be conceptualised as the systemic *emergent outcome* of a complex system. Complex systems in this context can be defined as systems composed of interacting adaptive agents. More precisely, the dynamics of a complex system are driven by the interaction taking place at the level of its components, and the components’ adaptations to the environmental changes they mutually generate. As such, social systems display some characteristic properties which do not lend themselves to reductionist approaches, but instead necessitate the adoption of ‘bottom-up’ models, i.e. models that generate aggregate patterns starting from an explicit representation of the behaviour of the components of the system – in this case, individuals; the direct interactions between the components; and interactions between the components and their environment.

Consequently, we propose that complex systems simulation methods, in particular various forms of agent-based modelling (ABM), are a critical component in the fight against the ‘wicked’ health problems facing us in the 21st century and should become a key part of the toolkit for tackling public health problems. These modelling techniques allow us to represent individual behaviours and their interactions, study the tangled web of causal relationships among environmental, physical and social factors affecting health-related behaviours, and simulate the effect of these relationships on the dynamics of public health problems at the population level.

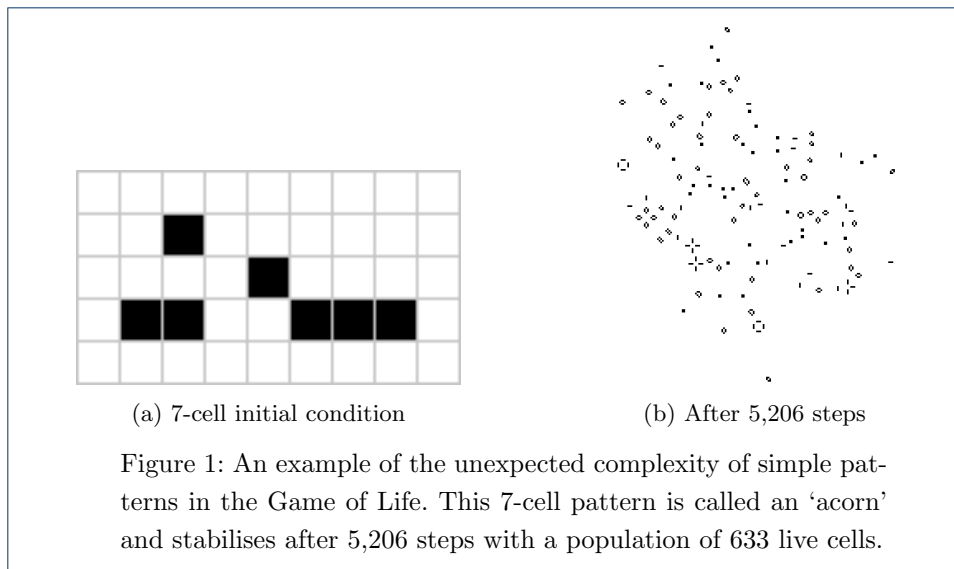
Ultimately, we take the view that decision-making related to complex, ‘wicked’ health problems must include the use of simulation in order to address these issues in a coherent way.

4 Challenges from complexity in public health

In line with our claim that public health is an emergent property of a complex system, in this section we explain the key characteristics of complex systems, the ways in which human society fulfills these characteristics, and the resultant impact on research efforts related to human social systems.

4.1 Emergence

A fundamental property of complex systems is *emergence*, which philosopher Mark Bedau divided into strong and weak forms in the context of complex systems [19]. Frequently, references to emergence in the natural sciences and philosophy are referring to *strong emergence*, which describes properties of systems that are not deducible from the behaviour of their component parts. To paraphrase Bedau’s example, the inscrutable phenomenon of consciousness is clearly a consequence of neural activity, yet our knowledge of the behaviour of neurons does not provide us with any insight into the function of consciousness [19]. Consciousness is not a property exhibited by individual neurons, and appears distinct from any particular neural property or behaviour, and still it arises from neural activity. Consciousness



can also change our neural activity, despite being distinct from it; this is *downward causation*, meaning that an emergent property can alter the behaviour of the component parts from which it emerges.

Strong emergence, while successfully capturing the idea of a macro-level property that is distinct from and yet capable of influencing its own components – as opposed to a macro-property that is merely an interesting consequence of micro-level activity – is philosophically problematic. Strongly emergent properties appear to be essentially autonomous from their components, and yet are able to exert strong causal influence on those same properties.

Consequently, Bedau’s exploration of *weak emergence* has become an important concept for complex systems science:

Macrostate P of S [system composed of micro-level components] with micro-dynamic [micro-level behaviour] D is weakly emergent iff [if and only if] P can be derived from D and S’s external conditions, but only by simulation (Bedau 1997 p. 4)

In other words, the behaviour of a system composed of interacting micro-level components is ultimately derived from its micro-level behaviours and the influence of its environment. If we can simulate these interactions explicitly, we can simulate the dynamic that generates the emergent property, and thus we can replicate the emergent macrostate of that system. Weakly-emergent properties are accordingly more accessible to scientific study, in that their emergence can be replicated via step-by-step simulation of the interaction of their constituting components and the surrounding environment. Conversely, if we do not simulate this dynamic, we cannot replicate the weakly-emergent behaviour.

This latter point is particularly important, as the macro-level emergent outcomes of a system’s microdynamic cannot be straightforwardly predicted, even with perfect knowledge of a system’s initial state and the rules driving its microdynamic. Bedau illustrates this using the Game of Life, a famous computational system in which cells on a grid change state according to the states of their neighbours. Cells in

Life have only two states – alive or dead – and their future states are determined by very simple rules according to how many of their neighbouring cells are alive or dead ^[1]. Despite this simplicity, even very simple starting configurations in Life can produce remarkably complex behaviour (see e.g. Figure 1), and Life can even play host to patterns capable of replicating any possible computation (a property known as *computational universality*) [20, 21]. As Bedau notes, this has profound implications:

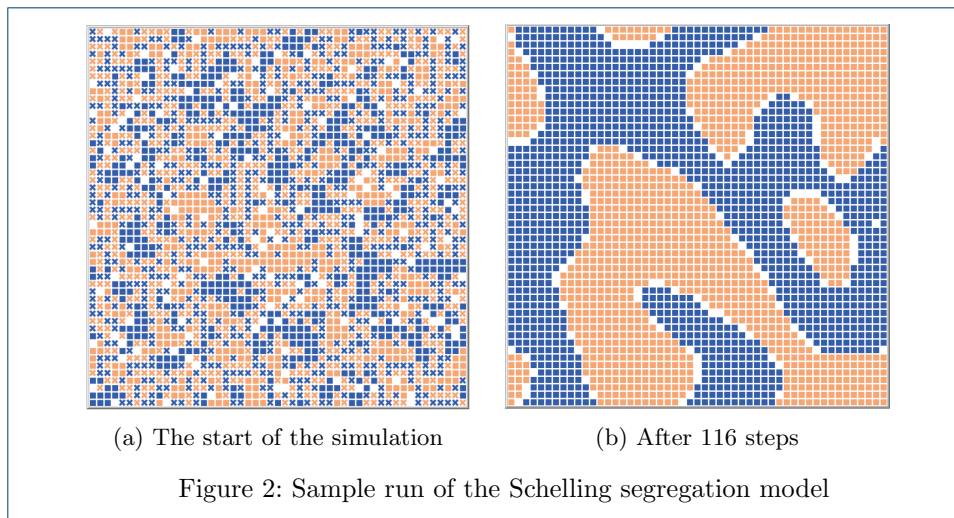
With few exceptions, it is impossible without simulation to derive the macrobehaviour of any state in a Life configuration even given complete knowledge of that configuration. In fact, since a universal Turing machine can be embedded in Life, the undecidability of the halting problem proves that in principle there can be no algorithm for determining whether the behaviour exhibited in an arbitrary Life world will ever stabilize. Yet all Life phenomena can be derived from the initial conditions and the birth-death rule. (Bedau 1997 p. 14)

Thus, the only way we can replicate the weakly-emergent macrostates of the Game of Life is to simulate its behaviour step-by-step. By extension, given that most complex systems will have significantly more complicated microdynamics than the Game of Life, replicating the macrostates of weakly-emergent systems requires the use of simulation to replicate their microdynamics.

A well-known example of an agent-based model replicating a weakly-emergent phenomenon from the interaction of micro-level entities is given by Schelling's residential segregation model [22, 23]. In this model, very simple agents are living in a virtual grid-based world, and at each discrete step of the simulation are able to choose to move their location. Their decision to move is based on a preference for the group composition of their neighbourhood; if the number of their neighbours belonging to a different group than themselves is above a certain threshold, the agent will move to a new random square on the grid. Here the agents' segregation is the emergent property of the system, while the preference for in-group neighbours is the parameter driving the agents' behaviour. Schelling showed that even a relatively low threshold generates a high degree of residential segregation, a result that is not predictable solely by knowing the agents' behavioural rules (see Figure 2). Thus we can describe Schelling's model as *weakly emergent*, given that its macrostate is derivable only by simulating the system step-by-step, despite its known and very simple behavioural rules.

Similarly, the phenomena of interest in public health research – from the development of health inequalities to the spread of obesity in certain communities – are consequences of complex interactions between individuals and their physical and social environments. Successful interventions may seek to influence the phenomena by altering individual, low-level behaviours through a number of different routes, in the hope that the population-level picture which emerges from those actions changes for the better – much like the simulation scientist tweaking the preferences

^[1]Life's rule is often expressed simply as B3S23: cells are (B)orn when they have exactly three live neighbours, and (S)urvive if they have two or three live neighbours. In all other circumstances, cells die or remain dead.



of Schelling’s agents in the hope of reducing segregation. Following Axelrod and Tesfatsion, we might align ourselves to using ABMs for normative understanding, or ‘evaluating whether designs proposed for social policies, institutions, or processes will result in socially desirable system performance over time’ [24].

If we accept that population-level health patterns are weakly-emergent phenomena deriving from the interactions between individual, society and environment, then it follows that answering some questions about those properties will require simulating those interactions explicitly via simulation. Agent-based modelling, as a methodology tailored to the investigation of emergent properties, is well-placed to provide insight into these phenomena by explicitly modelling both individual behaviours in response to an intervention and interactions between individuals and with their environment.

4.2 Non-linearity

In the context of complex systems science, there are two extant definitions of ‘non-linearity’. The first, which we will refer to as *causal non-linearity*, encompasses the manner in which complex systems tend to be characterised by cyclic, relational and mutual causal relationships between the variables describing the system’s state. This characteristic of complex systems means that popular causal inference methods like directed acyclic graphs (DAGs) cannot be used to characterise complex systems, as DAGs cannot include feedback loops.

The Schelling model again provides a useful example of this concept. Schelling’s simple agents change the local environment of both their previous and current neighbours, thereby affecting the probability that those neighbours may move to another location. In this way, the agents are affecting the actions of other agents indirectly via their shared environment.

Applying this concept in the context of public health, we might imagine an ABM that examines how unhealthy behaviours propagate in a social network by explicitly representing agents with their characteristics, social relations and interactions. For example, a simulation can explore the spread of unhealthy habits (such as smoking,

drinking, and drug use) among socially-related individuals, and evaluate the effectiveness of various strategies of network intervention that intend to induce desirable behavioural change across the social network.

Complex systems are characterised by a tangled web of non-linear causal relationships, which blurs the distinction between exogenous and endogenous variables. The non-linear causal relationship among the system's components and between the components and their environment is the fundamental reason behind our inability to forecast the system's dynamics from the components' behavioural rules: the web of causal relationships is simply too tangled for our limited cognitive capabilities to take all of them into account when trying to run a mental simulation of the system.

The second definition of 'non-linearity' refers to the kind of relationship between one or more exogenous variables and one endogenous variable: in this case, saying that the relationship is not linear means that variations in the exogenous and the endogenous variables are not proportional. In the Schelling model, we may change the agents' tolerance for different neighbours without noticing any significant change in residential segregation at the population level, as long as we are below or above a threshold level. Once we reach this level, however, a slight change in the agents' tolerance changes the system from a mixed state to a segregated one. Thus, the relationship between the agents' tolerance and the system's level of residential segregation is a non-linear one.

A similar relationship is found in the dynamics of infectious disease. The spread of an infectious disease is defined by the basic reproduction number, which is the average number of new cases caused by an infected individual during his infectious period. The basic reproduction number is a threshold that dictates whether the infection will persist over time. When the reproduction number is lower than one, the infection can not persist in the population, whereas if it is greater than one the infection will spread and persist. Non-linearity is significant to both the Schelling model and in infectious disease modelling, since it demonstrates how past trends may abruptly change once a certain threshold is crossed, producing a qualitative change in the state of the system: from mixed to segregated and from a non-persistent infection to an epidemic.

Of course, these two kinds of 'non-linearity' are strictly related. Complex systems are characterized by non-linear causal relationships between their components, and thus we often observe a non-linear relationship between exogenous and endogenous variables at the aggregate level.

4.3 Adaptive behaviour

Adaptive behaviour refers to the capacity or propensity for an agent to change its state following a change in its environment (including the behaviour of its neighbours)^[2]. This fundamental characteristic of complex systems allows for non-linear causality: an environmental variation prompts the agents' behavioural responses,

^[2]We should note here that adaptive behaviour is an area of study in its own right within complex systems, with a particular focus on embodiment and embodied cognition – the theory that the cognition and behaviour of living systems are shaped by the interaction between their physical form and the surrounding environment, not just the workings of the brain itself. In this framework, a complete description of the

which then feed back into additional environmental variations, and so on. In other words, the system's components may affect each other both directly and indirectly through changes in their common environment.

For example, consider the development of a new road that passes through a neighbourhood that will increase traffic, noise and air pollution in the area. As a result, residents who can afford to move may leave, and local housing prices may decline. The decline in housing prices may attract a less affluent population to the area. In this situation, individuals are adapting to new conditions in the environment, and the system – and thus the neighbourhood – self-organises as a result. In the new order that emerges an increasingly deprived population is located in a neighbourhood with poor environmental conditions and exposed to greater health risks.

Given that the components are the 'engines' of the co-evolutionary process driving the system's dynamics, the behavioural model of these components is the fundamental building block of any complex systems model. As these are weakly-emergent phenomena, we cannot replicate the dynamics of the system unless we simulate it as the result of the interaction between the system components and their environment.

Human societies are characterised by adaptive behaviour of the most complex kind, as human beings are able to recognise that they are in a complex system, identify the system's emergent properties and develop models that take them into account to drive their own actions. This phenomenon of *second-order emergence*, or the fact that emergent social institutions become part of the agents' models driving their behaviour, create direct causal relationships between the components' behaviour and the system's dynamics, which further compounds the complexity of the system [25].

4.4 The complex systems challenge to traditional epidemiology

Having outlined the defining characteristics of complex systems, we can better understand why they pose a challenge to the statistical approach typically adopted by epidemiology and how ABM can help epidemiology to rise to this challenge. Public health problems can be seen as the emergent outcomes of the complex social system that is human society. As such, to understand their dynamics we need to develop models based on the explicit representation of the components of society – individual human beings.

Individuals' adaptive behaviour and the resulting web of causal relationships between agents and their environment mean that non-linear relationships between system variables are pervasive in human social systems. This means that a very small variation of system inputs can generate a big variation in system outcomes, or vice versa. We can visualise a non-linear complex system as one where the space of possible outputs is very rough along the many input dimensions: because of the behaviour of an organism requires study not just of the organism itself, but the entire brain-body-environment system. This area of work has a strong computational tradition and is closely tied to work in complex systems, dynamical systems and artificial life. Here we use the term adaptive behaviour in a more general sense. For more on embodiment, the works of Francisco Varela *et al.* (*The Embodied Mind: Cognitive Science and Human Experience*) and Andy Clark (*Being There: Putting Brain, Body and World Together Again*) are excellent starting points.

number of factors affecting the relationship between any two variables, points that are very near to each other in the space of any input can be far apart in the space of outcomes. While the traditional statistical approach can be used *in principle* to shed light on the causal mechanism through which variable X affects variable Y (and in fact much epidemiological research consists of the addition of confounders and mediators to the original theoretical model to enhance our understanding of how variable X affects variable Y), success relies on the availability of a ‘sufficient’ number of observations for the analysis to have enough statistical power, a threshold that increases with the number of confounding variables in the causal model. This represents a limit to the complexity of the theoretical model that can be statistically analysed.

With respect to complex systems like human society, this creates two major problems. First, a complex system may contain variables for which it is difficult or impossible to gather empirical values. Second, even if our theoretical model does not contain such variables, in complex systems the number of *potentially* conditioning variables is typically very large, so we may have too few observations to conduct a meaningful statistical analysis of the relationship between the variables of interest, or reach the limits of analytic tractability of a mathematical model with dozens of variables.

Thus, we see the reason why most causal models in traditional epidemiology are relatively simple compared to ABMs: the number of observations must be large enough for the analysis to have the desired statistical power, while remaining analytically solvable. In other words, our tools force us to assume that numerous variables which we may ideally want to include in our models do not affect the relationship between X and Y. We call this the *stability assumption*, in that it requires that the relationship of interest is unaffected by changes in contextual variables.

Statistical approaches suffer further when data is sparse, as is often the case in human social systems. Properly-specified theoretical models can still be applied in these cases as means for increasing our understanding of system behaviour; such models can form the basis for the examination of ‘what if’ scenarios and for probing system behaviour via sensitivity analysis. We propose that ABM can be very effective in this regard, as the approach requires us to formally codify our theoretical knowledge of a system in the form of an explicit computational model of the processes underlying it. Through simulations we can produce counterfactuals, allowing us to evaluate which contextual variables we may exclude as conditioning variables, and whether the stability assumption is tenable. If the stability assumption does not hold, we can examine the effects of the conditioning variables on the relationship of interest. All the while we are able to model system processes explicitly, including non-linearities and feedback loops.

In this context, ABM can be seen as a complementary tool to assess the limits of statistical approaches as applied to a complex system, and to investigate system behaviour when quantitative data is too scarce to perform robust statistical analyses. While the traditional modelling of epidemiology is statistical - hypotheses relating to the causes of a health outcome are tested in a mathematical framework against observed data - the modelling approach for an ABM involves taking theories and assumptions underlying population health research and instantiating them in a computational framework.

5 Towards Model-Based Science

The change of modelling mindset required by agent-based modelling, and the related shift in emphasis toward theory and assumptions over quantitative data, may partially explain why the uptake of ABMs in population health is so minimal (systematic reviews dig up mere tens of papers, despite ABMs having been used to study human social systems for nearly 50 years [26]). These factors alone are not enough to explain the lack of engagement with ABM, however; other systemic and practical reasons contribute to the current state of affairs.

We suggest there are three main factors driving this lack of enthusiasm for the approach. Firstly, claims of what ABM can offer above traditional methods are contested. Newer statistical models may be capable of estimating causal effects in the presence of dynamic processes, treatment interference or spatial and network autocorrelation [10], thus removing some justifications for choosing ABMs over an alternative approach with which the community is already familiar.

Secondly, computer programming^[3], simulation modelling and complexity science are relatively uncommon skills in population health practice and research – unlike statistical theory, data management, and quantitative analysis. This is due to the traditional epidemiological focus on Positivist hypothesis-testing approaches to identify causal processes and intervention effectiveness. Adoption of ABMs, in contrast to novel statistical techniques which build on existing conceptual frameworks and skill sets, may require substantial retraining and changes in the way in which people think about and characterise population health challenges.

Finally, as mentioned above, there is a third crucial factor: ABMs within population health are frequently compared like-for-like with approaches designed to answer fundamentally different research questions. As a consequence, the methodology is perceived as imprecise and intractable in comparison to statistical methods, when in reality ABM is simply better-suited for different types of questions.

For example, a recent study compared ABMs and the parametric G formula in terms of how well they estimated the effect of a single treatment variable (anti-retroviral therapy) on a mortality outcome, when the decision to treat is dependent on two other variables [6]. Using ABMs for narrowly defined questions of causal inference is not the best use of the ABM method and, as we highlighted in our introduction, may not be the best sort of question to ask for complex population health challenges. Rather than identifying the causal effect of the treatment, alternative questions could be explored with ABM: what social factors could explain inequality in treatment adherence? What promotion strategies could facilitate faster uptake of treatment within communities?

In order to address these factors, population health researchers will need more exposure to ABMs applied to relevant problems, more opportunities to develop the skills needed to build simulations, and will need to apply their models to appropriate problems that leverage the strengths of ABMs while managing their weaknesses. We propose that a realignment of the discipline's approach to data and modelling

^[3]We refer to computer programming in the sense of using general-purpose programming languages, like C/C++, Python or Java, to construct self-contained computer programs. This is a distinct skill set from using domain-specific scripting languages, like R or Stata, to construct statistical models.

methodologies more generally will be needed, in order for ABMs to sit comfortably amongst the other methods in population health's toolbox.

5.1 The conceptual challenges of model-based science

Similar to other statistically-focused and empirically-inclined disciplines like demography [27], population health has progressed methodologically in a cumulative fashion. New statistical methodologies have spurred significant changes in research practices, yet each new methodology is fundamentally related to the previous, and no new method has completely replaced older ones. Each statistical tool still has its place in the metaphorical population health toolbox.

Agent-based modelling, however, does not fit this pattern as cleanly. ABMs are not a statistical methodology, and thus cannot simply be added to the methodological toolbox and applied to the same problems in the same way. If we take up multi-level modelling, for example, previous data can still be used and the methods themselves provide some guidance about what new data may prove useful. ABMs, however, have a more complex relationship with data, and in general what keeps the modeller awake at night is not a lack of data but instead the need for sensible parameter values for specifying agent behaviour.

So the newly-minted ABM researcher in population health must think not only about data sources, but some new topics as well:

- Social and behavioural theories: what should our agents be doing, and why?
- Uses beyond prediction: what can the ABM add to our knowledge of population change?
- Seeking out more diverse forms of data: can we alleviate our parameter problems using a wider variety of data sources, like qualitative data, GIS information, or participatory group model building?

This kind of shift in practice requires not only knowledge of a new approach and its epistemological and computational limitations, but a new perspective on how and when the approach should be used, and where it should fit in the overall scientific spectrum of population health.

Given the significant epistemological and practical implications of adopting agent-based methods, we propose that the relationship between ABMs and population health needs to change substantively before their usefulness becomes apparent. The development of a theory-based, model-centric approach to certain problems is needed, and in so doing we will uncover areas where ABMs are more suitable.

In our view, ABMs would best serve the interests of population health by being applied to areas where individual behaviour and interactions between individuals and their environments are influential in determining the success of interventions. In the next sections, we will discuss in more detail the potential role ABMs can play in a model-centric vision of population health, present some methods to mitigate the weaknesses of the approach, and finally illustrate some key areas of research where we believe ABMs can make a strong contribution.

6 Roles for Simulation in Population Health

Some of the most challenging 'wicked' problems in population health involve complex, interacting processes that are difficult to characterise in a traditional epidemiological framework. These challenges are likely to benefit from ABM research, given

that they can model the social and environmental aspects of population health explicitly.

However, this still begs the question: what would these models actually do? How would our quest to reduce or eliminate a ‘wicked’ problem benefit from an ABM, and how would it add to our knowledge? Here we illustrate a few ways in which ABM approaches can enhance our efforts.

6.1 Models as Policy Sandboxes

ABMs enable us to build a policy sandbox; a place to explore what evidence would be useful for taking certain decisions; to develop collaboratively a theory for why a certain policy will or will not work; and to test out how interventions could have an effect, if these theories hold true. Di Paolo, Noble and Bullock propose that while ABMs are often approached as a method to develop ‘realistic simulacra’ of the physical world, they may be best placed as platforms for the exploration of theoretical relationships within a system, their interactions and consequences [28]. Such conceptual exploration can be very useful for building theory. An approach that facilitates the apprehension of concepts in an abstract, rather than data-driven sense is useful for the precise reason that it differs from the data-driven approach of orthodox epidemiology.

In essence, traditional epidemiology focusses on risk factors and outcomes, with the link between them being an opaque black box [29]. However, in order to develop interventions to change the relationship between risks and outcomes, what is within that black box itself needs to be changed. This requires understanding mechanisms and having a theory of change in order to modify them. Statistical methodology does not provide the tools to model mechanisms, so researchers tend to focus on tasks that statistical methods can solve, which is the description of inequalities identified by the data without necessarily providing solutions. ABMs, in contrast, require people to model mechanisms explicitly, at least in the abstract, and provide a set of tools that encourages people to focus on the parts of the system where change might occur and could have the greatest impact. This leads us more toward solutions-focused research.

As Marshall notes, ABMs also provide a means to make use of a wider range of evidence:

Agent-based modeling represents one (but not the only) method to synthesize prior knowledge of a population – and the causal structures that act on this population – to understand how an intervention could affect the public’s health. In this manner, agent-based modeling is a science of evidence synthesis. Specifically, ABMs (and other simulation approaches) represent a platform for the integration of diverse evidence sources, including inconsistent or inconclusive scientific information, to support decision making for complex public health problems [30].

Further, the ability of ABMs to serve as theoretical exploration tools and intuition pumps when data is sparse can prove to be a significant advantage [31], as modellers would be able to investigate theories about population health using ABMs even in the absence of the expensive, difficult-to-collect data upon which

statistical approaches depend. Engaging in this kind of ‘model-based science’ can enable the evaluation of complex policy interventions amongst simulated populations, with outcomes serving as guides to decision-making under complexity rather than as point predictions of intervention outcomes.

There are additional pragmatic benefits to data-light, model-based investigations of interventions. Health data can be expensive to collect, as well as legally challenging due to strict data protection laws. Being able to circumvent these problems and use generative approaches to understand complex health problems would allow for potentially fruitful investigations of potential policy interventions when data is hard to obtain. Further, the outcomes of our simulations may help us identify where future data collection is needed.

6.2 Investigating Assumptions

ABMs have provided a method to run *in silico* interventions to inform policy makers about options, based on counterfactual model scenarios featuring agent populations using the same decision rules under varied environmental conditions. This scenario-based approach could be further augmented by allowing policy-makers direct input into which elements of proposed interventions are implemented, and the behavioural mechanisms underlying agent behaviours.

Traditional analysis of data is agnostic to the audience. ABMs in contrast provide a means to conduct post-normal science, or a scientific enterprise conducted in a context where ‘the puzzle-solving exercises of normal science (in the Kuhnian sense) which were so successfully extended from the laboratory to the conquest of Nature, are no longer appropriate for the resolution of policy issues of risk and the environment’ [32, p. 750]. Post-normal science acknowledges the pivotal role of input from the public and policy makers when dealing with environments suffused with unavoidable uncertainty. In this context, generative theories of causation still privilege scientific knowledge, but also integrate and adapt to the expertise of knowledge-users. The post-normal paradigm allows evidence to inform policies to improve health, moving beyond a context-free evidence approach.

Post-normal science does not yield to the opinions, values or politics of knowledge-users, and still adheres to the principles of calibration, validation and verification. The post-normal ABM approach provides opportunities for direct comparison of scenarios that include broader knowledge with those based solely on empirical data collection. Future research can examine how these approaches to model-building compare in their ability to inform health policy work and produce effective interventions.

From the perspective of post-normal science, ABMs function as a mechanism for reaching a collaborative agreement with knowledge-users about the emergence of population-level patterns in health from low-level interactions in social and spatial environments. The model provides feedback on the validity of the shared assumptions underpinning our understanding of critical population health issues. Exploring the model’s parameter space or implementing differing behavioural mechanisms enable us to compare different assumptions. In this paradigm, the process of model-building itself is an avenue for generating, developing, and comparing knowledge, and in the process, informing empirical and practical work on the phenomenon being modelled.

7 Understanding ABMs

As we have outlined, ABMs have numerous advantages when one wishes to explore the behaviour of a complex system. However, as with any methodology, ABMs are not a panacea and have aspects that can present significant challenges for the modeller.

One area in which ABMs remain opaque in some respects is in the analysis of their results. ABMs by their nature are suitable for modelling the behaviour and evolution of systems that defy formal statistical analysis, and that very complexity means they can appear opaque, where inputs and outputs are present but the influence of processes within the model are tangled and unclear. As a consequence, finding suitable parameter values for a simulation can be a lengthy process of trial and error, leading to significant time investment and frustration for the modellers.

This is an area where cutting-edge advancements in statistical and machine learning can work in concert with simulation modelling to produce insightful results. The growth of the field of uncertainty quantification and the current accessibility of machine learning opens up new possibilities for making ABMs more transparent and reducing these barriers to wider adoption.

7.1 Uncertainty quantification

The advent of the Bayesian Analysis of Computer Codes (BACCO) methodology has produced significant advancement in the analysis of model uncertainty and the impact of parameter values [33]. In particular, Gaussian Process Emulators (GPEs) have proven to be a promising method for analysing ABMs. In essence, a GPE takes a training set consisting of simulation outputs resulting from a wide range of input parameter values taken from across the parameter space, then develops a ‘surrogate model’ – a statistical model of the original simulation model [34]. The result is a summary of the fraction of the final output variance accounted for by each input parameter, as well as their interactions. This gives us a much clearer picture of the impact of each input parameter.

GPEs have been applied successfully in ABMs of various types, including models of social care [35], research funding allocation [36], and the effect of landscape changes on bird populations [37]. These analyses vastly increase the interpretability of ABM results, and provide much clearer guidance as to the function of each simulation parameter. Much of this work has been facilitated by the availability of GEM-SA, a free software package for analysing simulation results using GPEs [38]. As this technique continues to mature, and additional user-friendly means of implementation are created, variations of the GPE approach seem likely to become a common method for better understanding the internals of an ABM.

7.2 Machine-learning surrogate models

Machine learning, similar to GPEs, can also be used to develop ‘surrogate models’ which can allow for a detailed examination of the model’s behaviour and parameter space without an explosion of computational demands [39, 40]. Machine learning methods can improve the theoretical understanding of the ABM, help calibrate the model, and facilitate interpretation of the results relevant to end-users, therefore achieving an overall model of complex health issues with greater empirical validity.

Deep neural networks can also be used to create surrogate models, by training a multi-layered neural network on the output of numerous simulation runs [41]. Neural networks can approximate the output of any function, meaning that a trained neural network can approximate an ABM's output to a high degree of precision with much lower computational requirements. Neural-network surrogate models thus allow for much easier probing of an ABM's behaviour, and enable detailed sensitivity analyses to be performed. The surrogate model can also point us toward interesting areas of the parameter space that merit further exploration with the full simulation.

We should note that neural networks notoriously suffer from the issue of low interpretability, due to their numerous parameters, the complex structure of multiple hidden layers, and inherent non-linearity. However, we argue that interpretability is not a key requirement when using them as a surrogate of an ABM. If explaining the simulation result is of importance to the user, interpreting the ABM directly would be more suitable than explaining the prediction of the surrogate model. The trained neural network is intended to replicate the behaviour of the ABM to allow us to generate analyses of the model's behaviour more quickly. In this pragmatic role, our primary concern is that the neural network provides significant computational savings, not the particular way in which the network learns to replicate the ABM.

With the advent of these sophisticated methods for investigating the outputs of ABMs, we can begin to reduce their opacity and use them more effectively to understand how complex systems operate. Using surrogate models enables us to reduce an ABM's computational demands, while also facilitating deeper analyses of the model's underlying mechanisms. These analyses can then help us to understand how we can most effectively intervene at the micro-level of a population to produce positive changes at the macro-level.

8 Ways forward

As we have seen, ABMs and concepts drawn from the study of complexity could have a valuable role to play in future research in population health. The 'wicked' problems of the 21st century present a significant challenge to traditional statistical approaches, so approaching them from a complex systems perspective can enable us to characterise these problems in terms of the social and environmental interactions which lie at their core.

As with any field, not every tool is fit for every task, and this remains the case with ABMs and related methods. In the case of less 'wicked' issues in population health where causal links are more clearly defined, statistical methods already available and easily implemented can work quite well to generate the kind of effect estimates health researchers might seek. ABM should be a significant and powerful addition to the toolbox for population health modellers, used in harmony with the trusted epidemiological methods already in evidence.

For efforts extending further into explanation and description of more complex problems, ABMs start to shine. Below we describe three areas of potential interest population health researchers keen to experiment with ABMs, and note the difficulties facing traditional approaches in these contexts.

8.1 Health Inequalities

Health inequalities, which are systematic differences in the health status of different population groups, remain a difficult problem even in countries with otherwise high levels of equality [42]. For example, Nordic countries, despite generous welfare states and generally excellent population health statistics, still have persistent health inequalities [43]. Likewise, health inequalities have persisted in the United Kingdom even in the presence of free health care and systematic attempts by the New Labour government to reduce them [44].

A potential reason for this persistence is the sheer level of social reform that would be required to tackle this problem [44]. Wealth inequalities might need to be corrected, access to healthcare and hospital/clinic locations might need to be changed and optimised, and significant educational efforts to encourage changed health behaviours might need to be undertaken, amongst other changes. In the real world, policy interventions on this scale are not only cost-prohibitive, but quite disconcerting to policy makers, who are (perhaps understandably) reticent to restructure the whole of society on the basis of theories which in effect cannot be implemented without tremendous risk of unintended consequences [45]. These unintended consequences, or spillover effects, can potentially impact negatively on other key areas of population health [46].

ABMs, however, provide a testbed of sorts. Generative approaches and virtual populations can provide a risk-free laboratory for implementing large-scale policy interventions on simulated populations. While we would not wish to claim these kinds of models necessarily provide strong predictive power, if given sufficient detail in relevant social and behavioural mechanisms they could provide important insights on potential unintended consequences of these large-scale policy interventions.

More broadly, ABMs can shed light on those areas where health effects at the population level appear to be at odds with our expectations, given our knowledge of individual-level behaviours and processes. The case of inequality in Scandinavian welfare states above is a good example: we would expect very equal societies with universal access to free healthcare to show minimal health inequalities at the population level but, in reality, health inequalities are both still present and surprisingly significant. ABMs and complex systems approaches could help population health researchers unravel these mysteries by modelling the social and environmental effects that might lead to differential health outcomes despite society being set up specifically to avoid those outcomes.

8.2 Alcohol Use and Misuse

In the UK, the harm caused by alcohol both to the individual and to others is significantly higher than that caused by other drugs [47]. The role of psychological [48], social [49] and environmental factors [50] in influencing levels of substance use and dependence are increasingly recognised as key factors beyond biological addiction processes, both in terms of the risk processes, and for the most appropriate points of intervention. Scotland has recently implemented a minimum price per unit of alcohol [51].

Predicting the outcomes of population-level interventions is far from straightforward. Price-based measures are more likely to be effective amongst drinkers on the

lower end of the socioeconomic scale, whereas changes in pricing will have a greater impact on day-to-day spending habits [52], but the availability of alcohol (at any price) is strongly patterned by geography, with deprived areas of Glasgow having a greater density of alcohol, tobacco, fast food and gambling outlets, for example [53]. How alcohol retailers respond to minimum price legislation in areas of high compared to low supply may influence, and be influenced by changes in purchasing behaviour. Traditional statistical approaches are limited in the extent to which they can study non-linear interactions such as those between individuals and local retailers.

ABMs and complexity-inspired approaches can make substantive contributions in this area. Statistical modelling studies have looked in-depth at minimum unit pricing policies, but computational models could allow us to investigate how interactions between individuals, individuals and their neighbourhoods, and related effects such as socioeconomic differences [52]. Alcohol and drug abuse treatment outcomes are also affected by social factors [54, 55], and modelling these interactions and networks via ABMs could help in further fine-tuning interventions in this respect [56].

8.3 Obesity

Obesity is the target of substantial research and investment in population health worldwide, given that the condition is linked to such a wide range of health problems – both as an effect and a cause. Obesity stands out amongst the target research areas listed here, in that a number of simulation studies have already targeted obesity, and there have been some high-profile calls for systems-based modelling of the problem. The Institute of Medicine presented a book-length study of obesity prevention efforts in the United States, concluding that researchers and policy makers must take a ‘systems thinking’ perspective [57]. Chapter 4 provides a useful look at systems perspectives on the obesity epidemic – useful background for interested modellers [57]. Similarly, Foresight in the UK was commissioned to develop a strategic 40-year plan to tackle obesity in the UK, and presented a whole-system approach to the problem advocating large-scale societal, personal, and governmental changes [58].

Skinner *et al.* outline nine properties that make obesity well-suited as a target for systems-science-based modelling efforts, and present a systematic review of systems science models in obesity research ([59]). While their review finds dozens of studies using systems science perspectives and techniques, they conclude that the work done thus far has taken relatively limited views of the topic, investigating manageable portions of the overall obesity problem rather than tackling the larger, messier complex of societal and individual factors which drive obesity trends. Levy *et al.* take a similar view, noting that obesity modelling efforts are ‘at a nascent stage of development’ and just focus on ‘one or two links in the process’, leading the authors to advocate a comparative modelling approach [60, p. 390].

Obesity is thus a highly relevant area for modellers of a complex systems background, and indeed work of this type is already underway [61]. Much of the conceptual work around the social, behavioural and environmental factors influencing the obesity epidemic has already been done by respected sources. The widespread agreement that ambitious societal change at multiple levels will be required clears the way for similarly ambitious modelling projects to attempt to understand the

interplay of the multitudinous factors at play in obesity, and to develop productive collaborations with population health researchers and governmental bodies.

9 Conclusion

While population health research has contributed to numerous high-profile health successes in modern times, there remain some highly complex, ‘wicked’ problems which defy traditional methods of epidemiological analyses, and have resisted our attempts to develop effective interventions at the population level. These wicked issues in population health fall into the category of post-normal science – ‘These issues are urgent and of high public and political concern; the people involved hold strong positions based on their values, and the science is complex, incomplete and uncertain’ [62, p. 163]. Whereas traditional epidemiological methods are designed to reduce uncertainty about specific causal relationships, reliance purely on these methods inhibits the ability of the population health research community to say anything about complex wicked issues that urgently require solutions and where the scientific voice is non-existent or overwhelmed by caveats and uncertainty. Methods designed to illuminate complex interrelationships are unlikely to provide certain answers, but can provide an important contribution to the debate and may also provide a mechanism to bring together a range of alternative sources of knowledge.

As outlined above, there are numerous areas in which a complexity-inspired approach could contribute to efforts to develop more robust population health interventions for these wicked issues. These very brief examples demonstrate that developing interventions in these areas will require a broader, systems-based perspective due to the significant societal and behavioural change required. The complex systems science community is notable for its ability to apply computational approaches to an enormous range of domains, from artificial chemistry to simulated societies, by modelling the unifying mechanics and principles that describe the complex behaviour of these seemingly disparate systems. We suggest that this perspective makes this approach uniquely well-suited to undertake ambitious and challenging modelling projects aimed squarely at these wicked health problems.

Such efforts would require significant investment, not just in terms of time and finances, but in the form of interdisciplinary communication, which is no simple undertaking. Developing a common language between population health and complex systems will take time, but the potential benefit of efforts to enhance the health and wellbeing of millions of people seem worth the risk of occasional frustration and misunderstanding.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by UK Prevention Research Partnership MR/S037594/1, which is funded by the British Heart Foundation, Cancer Research UK, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Health and Social Care Research and Development Division (Welsh Government), Medical Research Council, National Institute for Health Research, Natural Environment Research Council, Public Health Agency (Northern Ireland), The Health Foundation and Wellcome.

MMcC holds a Medical Research Council/University fellowship supported by MRC partnership grant (MC/PC/13 027). ES is in the Complexity in Health programme supported by the Medical Research Council (MC.UU.00022/1) and the Chief Scientist Office (SPHSU16). MMcC is in the Relationships and Health programme supported by

Medical Research Council (MC_UU_00022/3) and Chief Scientist Office (SPHSU18). RS is supported by a MRC Mental Health Data Pathfinder Award (MC_PC_17217).

CA received funding from Research England's THYME Project, and from the Children's Liver Disease Foundation. These funders provided financial support for the authors, but were not involved in the development of this work or the preparation of this manuscript.

Author's contributions

ES, RS, and MM developed the concept for this paper and wrote the original drafts of the manuscript. UG, SP and JA significantly expanded the sections relating to complexity science core concepts. CA contributed the section on machine-learning surrogate models. All authors reviewed and edited the final manuscript in collaboration.

Author details

¹MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, 99 Berkeley Street, G3 7HR, Glasgow, UK. ²School of Computing, Engineering and Digital Technologies, Teesside University, TS1 3BX, Middlesbrough, UK.

References

1. Oliver, K., Lorenc, T., Innvær, S.: New directions in evidence-based policy research: a critical analysis of the literature. *Health Research Policy and Systems* **12**(1), 34 (2014). doi:[10.1186/1478-4505-12-34](https://doi.org/10.1186/1478-4505-12-34)
2. Krieger, N., Davey Smith, G.: The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology* **45**(6), 1787–1808 (2016). doi:[10.1093/ije/dyw114](https://doi.org/10.1093/ije/dyw114)
3. Vandembroucke, J.P., Broadbent, A., Pearce, N.: Causality and causal inference in epidemiology: the need for a pluralistic approach. *International Journal of Epidemiology* **45**(6), 1776–1786 (2016). doi:[10.1093/ije/dyv341](https://doi.org/10.1093/ije/dyv341)
4. Lipton, P.: *Inference to the Best Explanation*. Routledge, London (2003)
5. Marshall, B.D.L., Galea, S.: Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology. *American Journal of Epidemiology* **181**(2), 92–99 (2014). doi:[10.1093/aje/kwu274](https://doi.org/10.1093/aje/kwu274)
6. Murray, E.J., Robins, J.M., Seage, G.R., Freedberg, K.A., Hernán, M.A.: A Comparison of Agent-Based Models and the Parametric G-Formula for Causal Inference. *American Journal of Epidemiology* **186**(2), 131–142 (2017). doi:[10.1093/aje/kwx091](https://doi.org/10.1093/aje/kwx091)
7. Cherng, S.T., Tam, J., Christine, P.J., Meza, R.: Modelling the effects of e-cigarettes on smoking behaviour: implications for future adult smoking prevalence. *Epidemiology* **27**(6), 819–826 (2016). doi:[10.1097/EDE.0000000000000497](https://doi.org/10.1097/EDE.0000000000000497)
8. Schaefer, D.R., Adams, J., Haas, S.A.: Social networks and smoking: Exploring the effects of peer influence and smoker popularity through simulations. *Health Education & Behavior* **40**(1-suppl), 24–32 (2013). doi:[10.1177/1090198113493091](https://doi.org/10.1177/1090198113493091)
9. Auchincloss, A.H., Diez Roux, A.V.: A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *American Journal of Epidemiology* **168**(1), 1 (2008). doi:[10.1093/aje/kwn118](https://doi.org/10.1093/aje/kwn118)
10. Naimi, A.I.: Integrating complex systems thinking into epidemiologic research. *Epidemiology* **27**(6), 843–847 (2016). doi:[10.1097/EDE.0000000000000538](https://doi.org/10.1097/EDE.0000000000000538)
11. Yang, L., Gilbert, N.: Getting away from numbers: using qualitative observations in agent-based modelling. *Advances in Complex Systems* **11**(02), 175–185 (2008). doi:[10.1142/S0219525908001556](https://doi.org/10.1142/S0219525908001556)
12. Hiam, L., Harrison, D., McKee, M., Dorling, D.: Why is life expectancy in England and Wales 'stalling'? *Journal of Epidemiology & Community Health* **72**(5), 404–408 (2018). doi:[10.1136/jech-2017-210401](https://doi.org/10.1136/jech-2017-210401)
13. Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Diane Finegood, F.G., Harper, L., Hawe, P., Moore, L., Petticrew, M., Rehfuess, E., Shiell, A., Thomas, J., White, M.: Why we need a complex systems model of evidence for public health. *The Lancet* **390**(10112), 2602–2604 (2017)
14. Andersson, C., Törnberg, A., Törnberg, P.: Societal systems – complex or worse? *Futures* **63**, 145–157 (2014). doi:[10.1016/j.futures.2014.07.003](https://doi.org/10.1016/j.futures.2014.07.003)
15. Rittel, H.W.J., Webber, M.M.: Dilemmas in a general theory of planning. *Policy Sciences* **4**(2), 155–169 (1973). doi:[10.1007/BF01405730](https://doi.org/10.1007/BF01405730)
16. Signal, L.N., Walton, M.D., Ni Mhurchu, C., Maddison, R., Bowers, S.G., Carter, K.N., Gorton, D., Heta, C., Lanumata, T.S., McKerchar, C.W., O'Dea, D., Pearce, J.: Tackling 'wicked' health promotion problems: a new zealand case study. *Health Promotion International* **28**(1), 84–94 (2013). doi:[10.1093/heapro/das006](https://doi.org/10.1093/heapro/das006)
17. Academy of Medical Sciences: *Improving the Health of the Public by 2040: Optimising the Research Environment for a Healthier, Fairer Future*. Academy of Medical Sciences, London, UK (2016)
18. Galea, S., Riddle, M., Kaplan, G.A.: Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology* **39**(1), 97 (2010). doi:[10.1093/ije/dyp296](https://doi.org/10.1093/ije/dyp296)
19. Bedau, M.A.: Weak emergence. In: Tomberlin, J. (ed.) *Philosophical Perspectives: Mind, Causation, and World* vol. 11, pp. 375–399. Wiley, Hoboken, New Jersey (1997)
20. Gardner, M.: *Wheels, Life and Other Mathematical Amusements*. W.H. Freeman, New York (1983)
21. Rendell, P.: Turing universality of the game of life. In: *Collision-based Computing*, pp. 513–539. Springer, ??? (2002)
22. Schelling, T.C.: Dynamic models of segregation. *Journal of Mathematical Sociology* **1**, 143–186 (1971)
23. Schelling, T.C.: *Micromotives and Macrobehavior*. W. W. Norton, New York City (1978)
24. Axelrod, R., Tesfatsion, L.: A beginner's guide to agent-based modelling in the social sciences. In: Tesfatsion, L., Judd, K.L. (eds.) *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*, p. . North Holland, Amsterdam (2005)
25. Gilbert, N., Tierna, P.: How to build and use agent-based models in social science. *Mind and Society* **1**(1), 57–72 (2000)

26. Nianogo, R.A., Arah, O.A.: Agent-based modelling of noncommunicable diseases: a systematic review. *American Journal of Public Health* **105**(3), 20–31 (2015)
27. Courgeau, D., Bijak, J., Franck, R., Silverman, E.: Model-based demography: Towards a research agenda. In: Grow, A., Van Bavel, J. (eds.) *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, pp. 29–51. Springer, Cham (2017). doi:[10.1007/978-3-319-32283-4_2](https://doi.org/10.1007/978-3-319-32283-4_2)
28. Di Paolo, E.A., Noble, J., Bullock, S.: Simulation models as opaque thought experiments. In: *Seventh International Conference on Artificial Life*, pp. 497–506. MIT Press, Cambridge, MA (2000)
29. Susser, M., Susser, E.: Choosing a future for epidemiology: li. from black box to chinese boxes and eco-epidemiology. *American journal of public health* **86**(5), 674–677 (1996)
30. Marshall, B.: Agent-based modelling. In: El-Sayed, A.M., Galea, S. (eds.) *Systems Science and Population Health*, pp. 29–51. Oxford University Press, Oxford, UK (2017)
31. Silverman, E., Bijak, J., Noble, J.: Feeding the beast: Can computational demographic models free us from the tyranny of data? In: Lenaerts, T., Giacobini, M., Bersini, H., Bourguine, P., Dorigo, M., Doursat, R. (eds.) *Advances in Artificial Life, ECAL 2011*, pp. 747–754. MIT Press, Cambridge, MA (2011)
32. Funtowicz, S.O., Ravetz, J.R.: Science for the post-normal age. *Futures* **25**(7), 739–755 (1993). doi:[10.1016/0016-3287\(93\)90022-L](https://doi.org/10.1016/0016-3287(93)90022-L)
33. O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering and System Safety* **91**(10-11), 1290–1300 (2006)
34. Kennedy, M., O'Hagan, T.: Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B* **63**(3), 425–464 (2001)
35. Silverman, E., Hilton, J., Noble, J., Bijak, J.: Simulating the cost of social care in an ageing population. In: Rekdalsbakken, W., Bye, R.T., Zhang, H. (eds.) *Proceedings of the 27th European Conference on Modelling and Simulation*, pp. 689–695. Digitaldruck Pirrot, Dudweiler, Germany (2013)
36. Silverman, E., Geard, N., Noble, J.: Job insecurity in academic research employment: an agent-based model. In: *Proceedings of the Artificial Life Conference 2016*, pp. 460–467. MIT Press, Cambridge, MA (2016)
37. Parry, H.R., Topping, C.J., Kennedy, M.C., Boatman, N.D., Murray, A.W.A.: A bayesian sensitivity analysis applied to an agent-based model of bird population response to landscape change. *Environmental Modelling and Software* **45**, 104–115 (2013). doi:[10.1016/j.envsoft.2012.08.006](https://doi.org/10.1016/j.envsoft.2012.08.006). Thematic Issue on Spatial Agent-Based Models for Socio-Ecological Systems
38. Kennedy, M.: Description of the Gaussian process model used in GEM-SA. Software manual. <http://ctcd.group.shef.ac.uk/gem.html>. Accessed 23/05/2012 (2004)
39. Pereda, M., Santos, J.I., Galán, J.M.: A brief introduction to the use of machine learning techniques in the analysis of agent-based models. In: *Advances in Management Engineering*, pp. 179–186. Springer, Berlin (2017)
40. Lamperti, F., Roventini, A., Sani, A.: Agent-based model calibration using machine learning surrogates (2017)
41. van der Hoog, S.: Deep learning in (and of) agent-based models: A prospectus. arXiv preprint arXiv:1706.06302 (2017)
42. Bambra, C.: Health inequalities and welfare state regimes: theoretical insights on a public health 'puzzle'. *Journal of Epidemiology & Community Health* **65**(9), 740–745 (2011). doi:[10.1136/jech.2011.136333](https://doi.org/10.1136/jech.2011.136333)
43. Shaw, R.J., Benzeval, M., Popham, F.: To what extent do financial strain and labour force status explain social class inequalities in self-rated health? Analysis of 20 countries in the European Social Survey. *PLOS ONE* **9**(10), 110362 (2014). doi:[10.1371/journal.pone.01110362](https://doi.org/10.1371/journal.pone.01110362)
44. Mackenbach, J.P.: Can we reduce health inequalities? an analysis of the English strategy (1997–2010). *Journal of Epidemiology & Community Health* **65**(7), 568–575 (2011). doi:[10.1136/jech.2010.128280](https://doi.org/10.1136/jech.2010.128280)
45. Berkman, L.F.: Unintended consequences of social and economic policies for population health: towards a more intentional approach. *European Journal of Public Health* **21**(5), 547 (2011). doi:[10.1093/eurpub/ckr124](https://doi.org/10.1093/eurpub/ckr124)
46. Lorenc, T., Oliver, K.: Adverse effects of public health interventions: a conceptual framework. *Journal of Epidemiology & Community Health* **68**(3), 288–290 (2014). doi:[10.1136/jech-2013-203118](https://doi.org/10.1136/jech-2013-203118)
47. Nutt, D.J., King, L.A., Phillips, L.D.: Drug harms in the UK: a multicriteria decision analysis. *The Lancet* **376**(9752), 1558–1565 (2010)
48. Hammersley, R.: Constraint theory: A cognitive, motivational theory of dependence. *Addiction Research & Theory* **22**(1), 1–14 (2014). doi:[10.3109/16066359.2013.779678](https://doi.org/10.3109/16066359.2013.779678)
49. Sylvia, M.P., Alan, W., John, H.: All drinking is not equal: how a social practice theory lens could enhance public health research on alcohol and other health behaviours. *Addiction* **113**(2), 206–213 (2018). doi:[10.1111/add.13895](https://doi.org/10.1111/add.13895)
50. Birkmayer, J.D., Holder, H.D., George S. Yacoubian, J., Friend, K.B.: A general causal model to guide alcohol, tobacco, and illicit drug prevention: Assessing the research evidence. *Journal of Drug Education* **34**(2), 121–153 (2004)
51. Brennan, A., Meng, Y., Holmes, J., Hill-McManus, D., Meier, P.S.: Potential benefits of minimum unit pricing for alcohol versus a ban on below cost selling in England 2014: a modelling study. *BMJ* **349**, 5452 (2014). doi:[10.1136/bmj.g5452](https://doi.org/10.1136/bmj.g5452)
52. Holmes, J., Meng, Y., Meier, P.S., Brennan, A., Angus, C., Campbell-Burton, A., Guo, Y., Hill-McManus, D., Purshouse, R.C.: Effects of minimum unit pricing for alcohol on different income and socioeconomic groups: a modelling study. *The Lancet* **383**(9929), 1655–1664 (2014)
53. Macdonald, L., Olsen, J.R., Shortt, N.K., Ellaway, A.: Do 'environmental bads' such as alcohol, fast food, tobacco, and gambling outlets cluster and co-locate in more deprived areas in glasgow city, scotland? *Health & Place* **51**, 224–231 (2018). doi:[10.1016/j.healthplace.2018.04.008](https://doi.org/10.1016/j.healthplace.2018.04.008)
54. Schroeder, J.R., Latkin, C.A., Hoover, D.R., Curry, A.D., Knowlton, A.R., Celentano, D.D.: Illicit drug use in one's social network and in one's neighborhood predicts individual heroin and cocaine use. *Annals of epidemiology* **11**(6), 389–394 (2001)
55. Homish, G.G., Leonard, K.E.: The social network and alcohol use. *Journal of Studies on Alcohol and Drugs* **69**(6), 906–914 (2008). doi:[10.15288/jsad.2008.69.906](https://doi.org/10.15288/jsad.2008.69.906)
56. Apostolopoulos, Y., Lemke, M., Barry, A., Lich, K.H.: Moving alcohol prevention research forward – part I:

- introducing a complex systems paradigm. *Addiction* **113**(2), 353–362 (2018). doi:[10.1111/add.13955](https://doi.org/10.1111/add.13955)
57. Institute of Medicine: *Bridging the Evidence Gap in Obesity Prevention: A Framework to Inform Decision Making*. National Academy Press, Washington, D.C., USA (2010)
 58. Foresight: *Tackling Obesities: Future Choices*. Government Office for Science, London, UK (2007)
 59. Skinner, A.C., Foster, E.M.: Systems science and childhood obesity: a systematic review and new directions. *Journal of Obesity* **2013**, 10 (2013). doi:[10.1155/2013/129193](https://doi.org/10.1155/2013/129193)
 60. Levy, D.T., Mabry, P.L., Wang, Y.C., Gortmaker, S., Huang, T.T.-K., Marsh, T., Moodie, M., Swinburn, B.: Simulation models of obesity: a review of the literature and implications for research and policy. *Obesity Reviews* **12**(5), 378–394 (2011). doi:[10.1111/j.1467-789X.2010.00804.x](https://doi.org/10.1111/j.1467-789X.2010.00804.x)
 61. Silverman, E., Charles, F., Porteous, J., Wood, I., Ells, L.: Agent-based virtual urban environments for population health applications. In: *ABMUS 2017 : The 2nd International Workshop on Agent-Based Modelling of Urban Systems* (2017)
 62. Gluckman, P.: The art of science advice to government. *Nature* **507**, 163–165 (2014)