1. INTRODUCTION

The objective of this chapter is to convey the key concepts, overarching methods, and common applications of agent-based modeling in population health science. I aim to provide the reader with a foundational understanding of how and why agent-based models (ABMs) are increasingly employed to address pressing public health challenges of the 21st century. I will demonstrate these concepts with an example from my own work using an ABM to simulate HIV transmission dynamics in high-risk populations. Readers interested in gaining a more in-depth appreciation for the design, construction, and validation of ABMs are referred to several excellent texts on the subject [4, 5]. I conclude this chapter with a discussion of two promising avenues for the continued adoption of agent-based modeling approaches to improve population health: the evaluation of policy experiments and evidence synthesis.

1.1. What are Agent-Based Models?

ABMs are individual-based microsimulations that simulate the behaviors and interactions of autonomous “agents.” In most epidemiological applications, agents represent people who interact with each other to form an artificial society, thus simulating a hypothetical population of interest. However, an ABM can represent any discrete set of units that interact with each other (e.g., hospitals, schools, or governments).

The evolution of an ABM is determined by pre-programmed agent characteristics and by rules that regulate how agents behave, relate with each other, and interact with the simulated environment [7]. Even models with simple rules governing
agent behavior can result in complex, unanticipated population phenomena [8]. By comparing model output under different rule sets, hypothetical public health programs and policies can be implemented and tested. Interventions that change agent behavior, alter contact networks through which risk factors or diseases are transmitted, and/or modify environments in which health is produced can all be interrogated. Thus, ABMs serve as a highly flexible modeling laboratory in which a wide variety of interventions can be evaluated across populations and contexts. Although agent-based modeling within population health science has its roots in the study of infectious disease dynamics [9, 10], ABMs are increasingly used to explore the etiology and prevention of non-communicable diseases [11], social “contagions” (e.g., obesity, incarceration) [12–14], and the effects of place on health [15, 16].

1.2. Strengths and Challenges of Agent-Based Modeling

Like ant colonies, traffic jams, and stock markets, population health outcomes emerge from local interactions between autonomous units (i.e., people) and their environments [17]. Agent-based modeling, in which population-level phenomena arise as a result of micro-level interactions among the agents, is thus well suited to the study of many population health systems [18] (Box 8.1). In fact, the ability to simulate emergence—the appearance of larger entities, patterns, and regularities from interactions among smaller or simpler entities that themselves do not exhibit such properties—is a key strength of agent-based modeling approaches [19]. In systems epidemiology, the unique contribution of ABMs and other “bottom up” simulation tools stems from their capacity to reproduce (and thus understand) the processes through which group-level phenomena are generated. For this reason, ABMs are commonly used to elucidate the origins and determinants of population-level protective factors (e.g., herd immunity) and harms (e.g., clustering of disease susceptibility) in human populations [20–22].

A second strength of agent-based modeling techniques arises from the fact that agent-level outcomes are, by definition, non-independent. In most ABMs, an agent’s state is explicitly influenced by the status or behavior of other agents. Similarly, in infectious disease epidemiology and in many other health applications, an individual’s health status is influenced by the outcome(s) and/or exposure(s) of other people in the population. In contrast, many standard causal inference methods assume that the effect of the exposure on an individual is independent of the exposure of other individuals [23]. This assumption has been extensively characterized and is widely known as the stable unit treatment value assumption (SUTVA) [24]. The paucity of methods to identify causal effects in circumstances when SUTVA is violated has led to challenges in identifying the individual- and population-level effect of vaccines [25], and neighborhood-level
Box 8.1

HIV Transmission in High-risk Populations: An Agent-Based Model

Agent-based modeling has several distinct advantages over other mathematical modeling approaches. First, agents can take on a multitude of time-varying characteristics, which permits the simulation of heterogeneous, adaptive, and complex behavior in human populations. Second, agents interact with one another to form dynamic networks. Modeling these “relationships” permits the simulation and analysis of network effects (e.g., partner concurrency, epidemiologic bridging) that are difficult to capture with compartmental models [1]. We constructed an ABM to examine HIV transmission dynamics and the effectiveness of combination HIV prevention programs to reduce HIV incidence among high-risk populations [2, 3]. In this model, agents form sexual and drug-using relationships, which results in an evolving “risk network” through which HIV is transmitted. The model simulates an artificial society of 100,000 agents, representative of the sociodemographic characteristics, epidemiological profile, and risk behavior patterns of the adult population of New York City. This setting was originally selected because of the rich demographic, HIV surveillance, drug-using behavior, and social network data available. The model is coded in Python™ and simulates the passage of time in discrete monthly time steps.

Over the course of the simulation, partnerships are formed, maintained or broken, resulting in a dynamic sexual and drug-injecting network. The number of partners at each time step for each class of agent was specified by random sampling procedures from negative binomial distributions, using parameters from previously published estimates. Assortative mixing is incorporated by weighting the formation of partnerships between agents with similar characteristics. A snapshot of the agent network is shown in Figure 8.1.

In the ABM, the probability of HIV transmission depends on the infected agent’s HIV disease stage, HIV treatment status and adherence to therapy, the number of exposures (i.e., unprotected intercourse or syringe sharing) with an uninfected partner per time step, and the type of exposure (e.g., parenteral, sex between men, heterosexual). Moreover, interventions present in the agent environment modulate the likelihood of engaging in risk behavior and the probability that an HIV-diagnosed agent initiated antiretroviral therapy. For example, injection drug-using agents who are engaged with a needle and syringe program have fewer risk acts with their partners, compared with other agents.

First, the ABM was calibrated to reproduce empirically observed HIV epidemic trends observed in New York between 1992 and 2002. Then, HIV incidence trajectories from a “status quo” scenario (in which 2012 intervention coverage remains stable) were compared to those under hypothetical scenarios.
in which various HIV prevention interventions are scaled up. A “high-impact” combination prevention approach, in which coverage for all interventions was increased by 50% from 2012 levels, not only produced the lowest HIV incidence in 2040 but also resulted in a more immediate and sustained drop in new HIV cases. Importantly, HIV transmission was not eliminated under any scenario tested, demonstrating the importance of comprehensive, high-coverage interventions. More recent work has sought to determine whether HIV transmission occurring during early and acute HIV infection stages will hamper the success of antiretroviral-based prevention efforts [6].

**Figure 8.1** Network agent-based HIV transmission model. White: Agents acutely infected with HIV; Black: chronically infected agents; Gray: HIV-negative agent. Edges linking nodes represent past-month sexual and/or injecting risk behavior. Note the cluster of acutely-infected agents forming a “core” high-risk transmission group.
effects on health [26], among other areas. Agent-based models are not so constrained; disaggregating the influence of non-independence between units (i.e., “spillover effects”) from individual exposures (i.e., “direct effects”) is a useful and practical application of agent-based modeling in epidemiology [27].

Like all simulation models, the validity of ABMs depends on the strength and validity of assumptions made during the model development process. The availability (and accuracy) of empirical data with which to parameterize the model is also critical in many population health studies employing ABMs. The number of frameworks upon which to base the construction, calibration, and validation of ABMs continues to grow [28, 29]. Guidelines for reporting ABMs, including their key assumptions and more detailed elements, are also available [30, 31]. Nonetheless, agent-based modeling rests on the fundamental assumption that, for any given research question, people and their relevant health states, interactions, and environments can be credibly modeled at a reasonable level of abstraction [28]. Meeting this assumption carries at least two important challenges. First, the researcher must decide which minimal set of characteristics defining the agents, their relationships, and their environments are needed to accurately capture the processes being modeled. Second, data or prior knowledge must exist to inform the specific structure and parameterization of the model processes. In the sections that follow, we describe the most commonly employed methods to help ensure that an ABM has internal validity, and that the model’s results have relevance to solving real-world problems.

2. AGENT-BASED MODELING METHODOLOGIES

2.1. Implementation of agent-based models

For an ABM to run, the model developer must identify, define, and program agent behaviors, agent-agent interactions, and the relationship between agents and their environments. The first step in this process is to specify a “target”—the phenomena the model is intended to represent, reproduce, and simulate [4]. Since no model can capture all possible characteristics, behaviors, and environments that may influence a health outcome of interest, building a model for the target requires a theoretically motivated and conceptually grounded process of abstraction [29]. The goal is not to construct an all-encompassing representation of reality, but a highly simplified depiction that nonetheless provides valid insights into real-world phenomena and improves scientific understanding [32]. Commonly employed conceptual frameworks in population health, including the social-ecological model of health behavior and the syndemic theory of disease production [33, 34], can be helpful in identifying the core components of an ABM as employed in epidemiology. Given that simplification is a necessary step in all
models used in epidemiology, existing tools, including causal diagraming [35], may also be helpful in determining the key processes to be modeled in an ABM.

An ABM requires a set of code and a computational engine to be executed. Many ABMs are designed from scratch using all-purpose programming languages such as Python™, Java™ and C++. A number of modeling services (with preloaded libraries of commonly employed routines and functions) are also available. Among the most common programs are NetLogo (https://ccl.northwestern.edu/netlogo) and Repast (http://repast.sourceforge.net); other software has been reviewed elsewhere [28]. Once the model has been developed, the simulation is run by having agents repeatedly execute their behaviors and interactions according to pre-programmed rules. Most ABMs simulate the passage of time in steps that are either discrete or activity based.

In an ABM, agents are endowed with static or dynamic behaviors that can depend on everything from simple “if-then” rules to complex adaptive processes. The model developer must also define which agents are (or could be) connected to whom and the dynamics of these interactions. The way in which agents are connected is referred to as an ABM topology. Common typologies include a spatial grid, a more complicated spatial geography (e.g., agents can only interact with other agents who are nearby), or a social network. The topology of an ABM can also evolve over time. For example, in one study that used an ABM to evaluate policies to reduce influenza transmission in the workplace, agents were assigned to and moved between specific geographic locations representing schools or workplaces [36]. During each simulated day, agents could only interact with other agents who shared the same social activity location.

Agents can be entirely passive (i.e., purely responsive to exogenous stimuli) or seek to actively alter other agent characteristics or the environment. Therefore, unlike many socio-ecological models of health production which assume that health outcomes are determined by immutable upstream forces, ABMs allow individuals to not only interact with, but change, their environment(s). The environment may represent a set of geographic characteristics (e.g., pollutants, crime, or other aspects of the physical or built environment), venues (e.g., homes, bars, workplaces), or institutions (e.g., hospitals, prisons). An agent’s location in the simulated landscape is usually recorded as a dynamic attribute as they move in space. Different types of environments may promote, facilitate, or constrain agent behaviors. For example, in one ABM that simulated walking behaviors within a city, the agent environment was composed of 400 equal-sized neighborhoods, each with two properties: safety and aesthetics [37]. A walking index (representing each person’s walking experience) was a function of both individual agents’ characteristics (e.g., age, walking ability) and the safety and aesthetic quality of
all neighborhoods along a walking route. Each agent’s walking index in turn affected how much she/he would walk in the subsequent day.

2.2. Verification and Validation of Agent-Based Models

Model verification and validation are essential components of the ABM design process. Model verification asks the question, “Does the ABM work as intended?” while model validation seeks to determine whether the “right” model has been built. While there is no universal approach or consensus on how to conduct verification and validation in agent-based modeling, a number of principles and overall techniques are available [38–40].

The goal of verification is to determine whether the programming implementation of the conceptual model is correct. This process includes debugging, verification of all model calculations, and determining whether equations in the model are solved correctly. Some common verification methods include debugging procedures, code “walk-throughs” (a form of software peer review), and boundary testing (i.e., determining whether the model performs as expected at the parameter boundaries).

Broadly, model validation is a process through which the research team assesses the extent to which the model is a credible representation of the real world. Several techniques are available to demonstrate a model’s validity. For example, a model is said to have face validity if content experts believe the model behaves reasonably well by making subjective judgments about its performance. Most ABMs also go through a procedure known as empirical validation. Here, the developer determines whether the model is a good representation of the target by comparing whether model output is consistent with observed data. Model calibration refers to an iterative process through which unmeasured or poorly measured parameters are adjusted until the model output is in agreement with empirical data. A number of different approaches to conduct calibration and empirical validation have been developed and are reviewed elsewhere [41]. A variety of statistical tests can be used to determine if the model’s behavior has an acceptable level of consistency with observed data [42].

2.3. Interpretation of Results from Agent-Based Models

Interpreting the output of an ABM has distinct challenges from other mathematical modeling approaches. First, isolating the causal mechanisms and effects of one parameter can be difficult in models with a high degree of agent heterogeneity and many interdependent processes. For this reason, conducting parameter
sweeps over multiple variables (frequently simultaneously) to understand model behavior is recommended.

Second, ABMs can be highly sensitive to initial conditions. In complex and chaotic systems, even small perturbations in initial conditions can lead to large differences in model output. We have previously shown that the identification of causal effects is only possible in models that meet a regularity assumption called “ergodicity” (for any given set of initial conditions, the model output can be represented by an ensemble average as the number of model iterations increases) [27]. If non-ergodic behavior is present, the mean of model output across runs is not well defined, and casual effects comparing two counterfactual scenarios cannot be estimated.

Third, precise prediction of real-world phenomena under different inputs can be problematic when many assumptions are made regarding agent behavior and interactions. Given the complexity of many ABMs, the objective should not necessarily be to predict specific population outcomes under different scenarios per se, but to conduct a robust policy analysis, such that recommendations consist of an ensemble of policy options which perform well under plausible model specifications and are robust to model assumptions.

3. AGENT-BASED MODELS AND THE FUTURE OF EPIDEMIOLOGY

Epidemiology is, at its heart, a pragmatic discipline that identifies opportunities to control and prevent disease [43]. As such, epidemiologists must concern themselves not only with the isolation of causal effects but also with the identification of effective ways to intervene. Doing so necessitates more than an understanding of the causal mechanics that link an exposure with a health outcome. One must also consider the population context (i.e., the underlying prevalence of the health condition and related risk factors) and how multifactorial causal structures interact to produce disease [44].

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. Formalized methods and frameworks for the integration of diverse data streams into simulation models (and their implications for evidence-based
policy analysis) have recently been proposed for population health sciences [45, 46]. Continued adoption of these methods is warranted.

I wish to conclude this chapter by arguing that, although systems science in epidemiology and public health has “arrived,” more work is needed. Agent-based modeling of population health systems and the impact of hypothetical policy changes on health have led to novel scientific insights. However, the actual uptake of programs and policies informed by ABMs, and their subsequent evaluation in real-world settings, are lacking. A multidisciplinary and iterative science, in which model developers work collaboratively with interventionists and policymakers to implement, evaluate, and improve public health programs is needed. Further stymieing the uptake of agent-based modeling is the fact that systems science methods are not currently featured in much public health curricula or training [7]. Thus, the capacity for agent-based modeling to enhance the health of populations requires understanding and appreciation of the method among not just epidemiologists but also public health practitioners and population health scientists broadly.

4. ACKNOWLEDGMENTS

I wish to thank Dean Sandro Galea for his feedback and critical insights on early drafts of this chapter. This work was supported by an “Avenir” new investigator award from the National Institute on Drug Abuse (DP2-DA040236).

5. REFERENCES


