G Model
EPIDEM-231; No. of Pages 9
ARTICLE IN PRESS

Epidemics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Research paper

Defining epidemics in computer simulation models: How do definitions influence conclusions?

Carolyn Orbann[a,*], Lisa Sattenspiel[b], Erin Miller[b], Jessica Dimka[b]

[a] Department of Health Sciences, University of Missouri, Columbia, MO, 65211, USA
[b] Department of Anthropology, University of Missouri, Columbia, MO, 65211, USA

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form 11 August 2016
Accepted 11 December 2016
Available online xxx

Keywords:
Infectious disease modeling
Agent-based simulation
Infectious disease history
Epidemic criteria

ABSTRACT

Computer models have proven to be useful tools in studying epidemic disease in human populations. Such models are being used by a broader base of researchers, and it has become more important to ensure that descriptions of model construction and data analyses are clear and communicate important features of model structure. Papers describing computer models of infectious disease often lack a clear description of how the data are aggregated and whether or not non-epidemic runs are excluded from analyses. Given that there is no concrete quantitative definition of what constitutes an epidemic within the public health literature, each modeler must decide on a strategy for identifying epidemics during simulation runs. Here, an SEIR model was used to test the effects of how varying the cutoff for considering a run an epidemic changes potential interpretations of simulation outcomes. Varying the cutoff from 0% to 15% of the model population ever infected with the illness generated significant differences in numbers of dead and timing variables. These results are important for those who use models to form public health policy, in which questions of timing or implementation of interventions might be answered using findings from computer simulation models.

1. Introduction

Computer simulation models are well-established tools within epidemiology and related disciplines. They have proven useful, powerful, and accessible for evaluating questions about the spread of infectious disease within and between many populations and over physical and social spaces (see, for example, Ferguson et al., 2005; Lofgren et al., 2014; Longini et al., 2005; Mahmoud et al., 2006; Yang et al., 2011). Computer simulation models are now employed by a wider range of researchers than ever before, likely due to changes in computing power, availability of software, and accessibility of platforms. While the increased use of models has made positive contributions to an ever-widening group of fields, cross-disciplinary communication about mechanics of modeling can be difficult and may lead to obstacles at any step of the modeling process.

This study begun after the authors noted that many disease modeling papers are unclear in reporting a central aspect of model outcomes: whether an epidemic has or has not happened. Very few modeling papers explicitly address the specific definitions and/or case thresholds used to characterize a simulation run as an epidemic within the model environment. Communication about models is important if replication is to be achieved and if policymakers are to use information from models in order to build effective public policy. As the creation and use of agent-based models proliferates, a discussion about the need for standardized methodology in model creation and reporting has followed. Some have proposed standards for describing model structure, such as the ODD protocol (Grimm et al., 2006; Grimm et al., 2010). However, these protocols do not address disease models specifically, but are directed towards agent-based models in general.

We focus here on how decisions made during the development and analysis of agent-based computer simulations of infectious disease transmission can influence the conclusions drawn from these models. Specifically, we center on how the criterion used to define whether a disease outbreak should be considered an epidemic affects inferences derived from the analysis of simulation outcomes. Because of the stochastic nature of agent-based simulations, most analyses of such models start by averaging the results from large number of runs of the simulation. However, even in models for highly infectious diseases, which *usually* infect large numbers of agents, occasionally the infection may result in only a small number of cases. If small outbreaks are relatively common

* Corresponding author at: Department of Health Sciences, 501 Clark Hall, University of Missouri, Columbia, MO, 65211, USA.
E-mail addresses: orbannc@health.missouri.edu (C. Orbann), sattenspiell@missouri.edu (L. Sattenspiel), elmvw9@mail.missouri.edu (E. Miller), DimkaJ@missouri.edu (J. Dimka).

http://dx.doi.org/10.1016/j.epidem.2016.12.001
1755-4365/

(but still not the norm), including them in analyses to identify general epidemic patterns may skew the analyses. Yet, there are few guidelines to use in determining whether a small outbreak should be identified as an epidemic run and therefore included in the calculation of averages across runs or whether it should be omitted from the calculations.

In order to systematically explore the impact of variable criteria in assigning epidemic status to a simulation run, we use different epidemic definitions in analyses of results from simulations of an agent-based model for the spread of communicable diseases. We specifically examine the impact of variation in the criterion used to define when an epidemic occurs on the interpretation of outcomes such as epidemic size, timing, and severity. For example, if 1% of the model population is required to become infected in order to include a run in an analysis, how does that compare with the results from a similar analysis in which a 5% threshold is used? How do these kinds of decisions impact the way in which public health efforts might be designed or implemented?

2. Defining "Epidemic"

A survey of the epidemiological literature has revealed that epidemics are loosely defined in a way that is broadly adaptable to a wide variety of conditions. The term has a long history and is used by practitioners in many health-related domains, each domain having their own nuanced understanding of what an epidemic is and when it is occurring. Within the disease modeling literature, the term is treated as uncontroversial, and at least for the last several decades, if it has been defined at all, a usual definition includes reference to a basic reproduction number (R_0) greater than one (Addy et al., 1991; Lee et al., 2010), which guarantees that one or more secondary and later generations of cases are spawned from the index case in the model. Using this broad definition, most runs of a model could be considered epidemics and thus included in data analysis. Analysis may then focus on grouping runs with similar timing patterns or geographic spread of the epidemic (Ajelli et al., 2010; Rahmandad and Sterman, 2008).

Currently, the Centers for Disease Control (CDC) uses two words to describe illnesses within populations that occur over a specific period of time, epidemic and outbreak. The first, **"**epidemic", is defined as

> "The occurrence of more cases of disease, injury or other health condition than expected in a given area or among a specific group of persons during a particular period. Usually, the cases are presumed to have a common cause or to be related to one another in some way" (Centers for Disease Control and Prevention, 2012).

The definition of "outbreak" is almost identical to that of epidemic. The only distinction is that an outbreak is "Sometimes distinguished from an epidemic as more localized, or the term less likely to evoke public panic" (Centers for Disease Control and Prevention, 2012).

Three important elements appear in both of these definitions. First, epidemics and outbreaks are both bounded by geography or population. Epidemics may occur in groups of people who do not necessarily share geography, but share other common attributes, such as occupation, cultural sub-group membership, or other uniting behavioral features. The second is the inclusion of a chronological element. Outbreaks and epidemics take place during a specific period of time and are defined in reference to a previous period of time. Finally, epidemics and outbreaks are caused by something, but that something is not necessarily a pathogen (this proves to be a particularly contentious element of the definition and will be further discussed below).

Given the similarity in their definitions, what makes an epidemic distinct from an outbreak? Popular usage of the terms plays a role and the term "epidemic" carries connotations that cause the public to react with concern or fear in a way that the term "outbreak" does not, at least at the current time. Disease modelers use both terms in a way that seems intuitive (i.e. outbreaks are small clusters of disease, whereas epidemics are more pervasive), but again, the definitions are not explicit and thus rely on the authors and the reader sharing the same assumptions about the terms and their usage.

The earliest use of the term epidemic in reference to medical phenomena can be traced to Hippocrates (Martin and Martin-Granel, 2006), who used the word "*epidemos*" to describe both a group of cases of a single illness and cases of multiple illnesses within a population over a specific period of time. Hippocrates' use of the term to describe illness within a population distinguished it from the common usage of *nosos*, a word used to describe illness within an individual. *Epidemos* was in use prior to Hippocrates, but not typically used in a medical or health-related way, instead referring to other kinds of phenomena that might be affecting a population, like famine or political strife. Martin and Martin-Granel (2006) provide a discussion of the previous uses of the term and how it evolved into modern usage.

At the time of Hippocrates, theories of disease causation were different than the prevailing theory in place today. Then, disease was thought to result from imbalances in natural forces, environmental factors like air and water, personal conflict, or other features. The current prevailing theory of disease causation, germ theory, centers on contagious items (contagions) that can be passed between individuals (Lederberg, 2000). The widespread adoption of germ theory resulted from discoveries in bacteriology and increases in microscopy technology during the 19th and early 20th centuries (Pelting, 2013). The mismatch between these different conceptions of disease causality means that it is not possible to say whether early medical practitioners thought of *epidemos* as applying only to what we now term "infectious disease".

Martin and Martin-Granel (2006) attribute the modern understanding of the term to changes in usage during the bubonic plague outbreaks of the Middle Ages in Europe. Bubonic plague has distinctive symptoms and was easily identified by lay people. Epidemics of plague were recognized quickly and known to result from a single disease, as opposed to epidemics in which the visible symptoms were more varied and thus the origin of the illness unknown. Thus, people began to associate the term "plague" with the epidemic or pandemic spread of a single disease. By the time the field of epidemiology developed in the 19th and early 20th centuries, epidemics were commonly understood as resulting from infectious diseases, like smallpox, measles, thrush, and dysentery among others, though the exact cause of each disease was unclear (Creighton, 1894; Crookshank, 1920; Farr, 1840; Parkin, 1873). At the same time, some still perceived connections to weather, magnetism, or electric fields as important factors in disease spread (Parkin, 1873). Epidemiological research at the time focused on trying to identify and characterize any natural laws of epidemics, using statistical and mathematical methods (Brownlee, 1915; Fine, 1979; Serfling, 1952)

In the late 20th and early 21st centuries, the term "epidemic" has come to be applied to any number of conditions or phenomena that spread within a population or over a geographic area. In addition to infectious diseases, people now recognize epidemics of non-communicable diseases and conditions (e.g. osteoporosis (Fogelman and Ryan, 1990), obesity (World Health Organization, 2000), diabetes (Gambert and Pinkstaff, 2006), and autism (Wazana et al., 2007)), epidemics of behaviors and accidents (e.g. texting and driving (Atchley and Geana, 2013), falls (Mayer et al., 2006), and prescription drug abuse (Paulozzi et al., 2012)), and epidemics of

Please cite this article in press as: Orbann, C., et al., Defining epidemics in computer simulation models: How do definitions influence conclusions? Epidemics (2016), http://dx.doi.org/10.1016/j.epidem.2016.12.001