Epidemics 10 (2015) 78–82

Contents lists available at ScienceDirect

Epidemics

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

Seven challenges for model-driven data collection in experimental and observational studies

J. Lessler[a,*], W.J. Edmunds[b], M.E. Halloran[c,d], T.D. Hollingsworth[e,f,g], A.L. Lloyd[h,i]

[a] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21224, USA
[b] London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
[c] Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA 98195 USA
[d] Center for Statistics and Quantitative Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[e] Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK
[f] School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK
[g] Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK
[h] Department of Mathematics and Biomathematics Graduate Program, North Carolina State University, Raleigh, NC 27695, USA
[i] Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 1 March 2014
Received in revised form 5 December 2014
Accepted 8 December 2014
Available online 16 December 2014

Keywords:
Modeling
Data collection
Observational studies
Experimental studies

ABSTRACT

Infectious disease models are both concise statements of hypotheses and powerful techniques for creating tools from hypotheses and theories. As such, they have tremendous potential for guiding data collection in experimental and observational studies, leading to more efficient testing of hypotheses and more robust study designs. In numerous instances, infectious disease models have played a key role in informing data collection, including the Garki project studying malaria, the response to the 2009 pandemic of H1N1 influenza in the United Kingdom and studies of T-cell immunodynamics in mammals. However, such synergies remain the exception rather than the rule; and a close marriage of dynamic modeling and empirical data collection is far from the norm in infectious disease research. Overcoming the challenges to using models to inform data collection has the potential to accelerate innovation and to improve practice in how we deal with infectious disease threats.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

Introduction: What is the role of models in data collection?

When people refer to "models" of infectious disease transmission, they usually mean something far more specific than the word "model" indicates. The term is generally used to refer to a system of equations or computer program that explicitly represents the mechanisms of disease transmission and pathogenesis. This is in contrast to models purely of statistical association that are common throughout the medical literature. By setting forth a mechanistic hypothesis in a mathematically precise form, models become tools for generating (perhaps unexpected) predictions which can be used to test the underlying hypotheses through confrontation with data. Though this use of models has a long tradition throughout all branches of science (including infectious disease epidemiology and ecology), it often takes a back seat to other uses of infectious disease models. The highest profile infectious disease modeling work is often aimed at making predictions or filling in gaps in existing data. These uses of models by definition *presume that the hypothesis captured in the model is close enough to the truth to capture the dynamics of the system relevant to the task at hand*, and provide answers only conditional on the correctness of the model. While many researchers put great effort into fitting both the structure and parameters of models using existing data, data are rarely collected with the explicit purpose of testing model hypotheses, and many models go unchallenged and untested after they are first presented.

Models are powerful, in part, because they can turn a hypothesis or theory into a tool for making precise predictions. Yet even in this capacity infectious disease models are underutilized in the data collection process. For example, sample size calculations and power analyses are de rigueur for the design of observational studies and clinical trials. However, cases of transmissible infections are non-independent, limiting the utility of standard theory. Using mechanistic models that account for the transmission process can allow robust estimation in the setting of these "dependent happenings" and tell us not only how much data to collect but when to collect it (Halloran et al., 2010).

* Corresponding author. Tel.: +1 410 955 3551.
E-mail address: justin@jhu.edu (J. Lessler).

http://dx.doi.org/10.1016/j.epidem.2014.12.002
1755-4365/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

Model-driven data collection does occur, and a prime example is the Garki project (Molineaux et al., 1980). This seminal project studied the transmission of malaria under different control interventions and captured longitudinal data on both disease and vector dynamics. Thirty years after its publication it remains one of the most valuable datasets for parameterizing malaria models (e.g., Griffin et al., 2010), and has been cited over 650 times. In part because input from models was used in the design of the study, it provides an almost unique dataset on the dynamics of malaria across a number of seasons and under different interventions, providing an example for future studies which has rarely been emulated.

Through sensitivity analysis and quantification of uncertainty, models can be used to elucidate the parameters and processes that contribute most to our inability to make predictions with a high degree of confidence, and those which are less important. They can thereby indicate where experimental effort would be best directed to improve the predictive power of a model. An example of where this is being attempted is HPTN 071 (PopART), a community randomized trial of combination prevention packages including early antiretroviral therapy initiation to control HIV transmission in Africa. In the analysis of this trial, models are being used to estimate endpoints (e.g., community incidence), improve study design, and identify which processes are the main drivers of uncertainty (Cori et al., 2014).

For the modeling of infectious disease to reach its full potential, it must become more tightly integrated with the data collection process. Despite successes such as those described above, significant challenges to successful model-driven data collection remain. Some stem from technical or cultural issues that are, in principle, easily surmountable. However, others are inherent in the role that models are often called on to play in public health response and scientific research, hence may be difficult or impossible to overcome.

In this paper, we start by discussing some fundamental challenges in the relationship between models and data collection, and end with specific challenges in current practice. Throughout we highlight the potential role of "modelers" in the data collection process. Though infectious disease modeling has grown into a distinct specialty, here we refer to anyone who defines and implements a mechanistic model of disease transmission as a "modeler". In addition to our headline challenges, we have tried to identify specific research avenues under each (mentioned at the end of key paragraphs) that could result in significant progress in confronting the challenge.

1. Ensuring a strong empirical basis for models used to fill gaps in data and knowledge

One primary role of models is to use our mechanistic understanding of a system to fill gaps in available data. Gaps in knowledge may occur because data is difficult or expensive to collect, because we are trying to understand past events which can no longer be directly observed, or because we are confronting a novel disease which is not yet well understood. In each of these contexts models play a hugely important role because of the lack of information, but are at the same time hamstrung by the lack of data with which to test model assumptions or fit parameters. While data will always be scarce in these situations, each presents an opportunity to use mechanistic models to make more effective use of existing data and guide ongoing data collection.

Data on disease incidence is difficult and expensive to collect on a broad scale; and passive clinical surveillance may include only the most severe cases or be clouded by a non-specific clinical profile. However, the burden of disease is one of the most fundamental pieces of epidemiologic information used in setting the public health agenda. Hence, mechanistic models are often used to fill the gap. For instance, transmission models have been used to help estimate the global burden of measles (Simons et al., 2012), using our knowledge of how susceptibility drives epidemic dynamics to infer the true number of cases from what was observed. Likewise, models have been used to help translate observed cases of acute flaccid paralysis to polio incidence through our understanding of transmission and the symptomatic attack rate (Eichner and Dietz, 1996). There are opportunities to test and improve models that fill data gaps. Models can be used to identify efficient data collection activities that validate the model but do not require the effort and expense of collecting the data the model is meant to infer. Innovative methods and systems for updating both the model predictions and the model assumptions in real time as new surveillance data becomes available could greatly increase their value in public health practice.

Understanding disease dynamics is often dependent on measuring disease incidence years or decades in the past, making the design of suitable data collection particularly challenging. Past incidence rates may be unmeasured because a disease circulated before it was identified (e.g., HIV before the 1980s), because acute infection often occurs without identifiable symptoms (e.g., dengue, HIV), or because of poor surveillance. In the latter two cases, even recent incidence patterns may be unknown. Dynamic models can often be used to infer past incidence with current cross sectional data or historic samples. Age specific serologies can be used to estimate the past force of infection, and have been used to measure historic patterns of the force of infection for dengue and other diseases (Rodríguez-Barraquer et al., 2013). Phylogenetic models can be paired with simple epidemic models to infer past epidemic dynamics, as has been done with HIV and hepatitis C (Stadler et al., 2012). These techniques, particularly phylogenetic inference, have become quite popular, but validation has been largely limited to simulation studies (e.g., Robinson et al., 2013; Volz et al., 2012). Studies aimed at collecting prospective data specifically to evaluate serologic and phylogenetic approaches to inferring incidence would help to place these inferences on firmer footing.

When responding to emerging epidemics, the problem of missing data is particularly acute. Here we are forced to forecast the course of an epidemic with limited knowledge of the pathogen and burden of disease. Critical data must be collected to carry out this task, some of which can be measured most effectively early on in the process of disease emergence. However, this data is not routinely collected early on, whether due to the difficulty of collection, competing priorities or its value being unrecognized. In particular, the tendency is to focus almost exclusively on cases early in an epidemic, whereas those who were at risk but did not become infected may carry the most information in terms of population susceptibility and disease transmissibility. A notable exception is the 2009 H1N1 influenza pandemic in the United Kingdom, where data collection was guided by the long-term involvement of modelers in the design of control programs. Although not all the data requested was collected, these efforts enabled policy-relevant modeling during the early stages of the epidemic (Ghani et al., 2010; Baguelin et al., 2010; Eames et al., 2012). This experience illustrates how integration with the public community can pay off in better inferences to support policy. At the time of writing, modeling is playing an important role in the response to the Ebola outbreak in West Africa, making use of the detailed contact tracing data collected as part of the response (though for purposes other than modeling) (WHO Ebola Response Team, 2014). Researchers should challenge themselves to identify the most useful classes of models in an emerging epidemic and the data needed to parameterize them. They should then work with public health officials to integrate collecting this data into epidemic response plans before such a response is needed.

Download English Version:

https://daneshyari.com/en/article/2813512

Download Persian Version:

https://daneshyari.com/article/2813512

Daneshyari.com