Epidemics 6 (2014) 18–29

Contents lists available at ScienceDirect

Epidemics

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

Fitting outbreak models to data from many small norovirus outbreaks

Eamon B. O'Dea[a,*], Kim M. Pepin[b,c], Ben A. Lopman[d,1], Claus O. Wilke[a,e]

[a] Section of Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA
[b] Fogarty International Center, NIH, Bethesda, MD 20892, USA
[c] Department of Biology, Colorado State University, Fort Collins, CO 80523, USA
[d] Gastrointestinal, Emerging and Zoonotic Infections Department, Centre for Infections, Health Protection Agency, London NW9 5EQ, UK
[e] Center for Computational Biology and Bioinformatics and Institute for Cell and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received 11 June 2012
Received in revised form 23 October 2013
Accepted 23 December 2013
Available online 8 January 2014

Keywords:
Stochastic epidemic model
Parameter estimation
Norovirus
Health-care-associated infection
Generalized linear model

ABSTRACT

Infectious disease often occurs in small, independent outbreaks in populations with varying characteristics. Each outbreak by itself may provide too little information for accurate estimation of epidemic model parameters. Here we show that using standard stochastic epidemic models for each outbreak and allowing parameters to vary between outbreaks according to a linear predictor leads to a generalized linear model that accurately estimates parameters from many small and diverse outbreaks. By estimating initial growth rates in addition to transmission rates, we are able to characterize variation in numbers of initially susceptible individuals or contact patterns between outbreaks. With simulation, we find that the estimates are fairly robust to the data being collected at discrete intervals and imputation of about half of all infectious periods. We apply the method by fitting data from 75 norovirus outbreaks in health-care settings. Our baseline regression estimates are 0.0037 transmissions per infective-susceptible day, an initial growth rate of 0.27 transmissions per infective day, and a symptomatic period of 3.35 days. Outbreaks in long-term-care facilities had significantly higher transmission and initial growth rates than outbreaks in hospitals.

© 2014 The Authors. Published by Elsevier B.V. Open access under the CC BY-NC-SA license.

Introduction

A common and difficult problem in epidemiology is to estimate rates of disease spread. Accurate estimates of these and other population parameters are crucial in the evaluation of disease control measures (Anderson and May, 1992; Keeling, 2005; Halloran et al., 2009) or biological hypotheses (Lively, 2010). Heterogeneity complicates the problem of obtaining such estimates. For example, a person's risk of infection depends on contact rates and acquired immunity, and these quantities can vary widely between people and outbreaks.

Norovirus (NoV) epidemiology provides a fine case in point of the need for models to accommodate heterogeneity. Noroviruses are the most common cause of diarrheal disease in the United States, causing an estimated 21 million cases (Scallan et al., 2011) and 71,000 hospitalizations per year (Lopman et al., 2011). A genetically diverse group of strains is often circulating within a population. New strains of the predominant genogroup 2 genotype 4 (GII.4) taxon appear regularly over time (Glass et al., 2009), and a person's risk of infection, given exposure, likely depends on both the antigenicity of the virus and the type-specific immunity developed from the person's previous exposure (Cannon et al., 2009). Other important heterogeneities include innate susceptibility (which depends on a person's histo-blood group antigens and secretor status) and age-specific risks of exposure. Outbreak investigations (Evans et al., 2002; Thornley et al., 2011; Wikswo et al., 2011) have provided convincing evidence that single vomiting incidents in crowded settings can lead to scores of secondary cases. Models that account for both between-individual and between-population heterogeneity are needed to obtain the accurate parameter estimates required for predicting outbreak dynamics and implementing effective controls. At present, control measures are based on general infection-control principles (Centers for Disease Control and Prevention, 2011) and thus are likely to be somewhat inefficient.

A further complication for modeling norovirus transmission is that it often occurs in small outbreaks. The transmission and

* Corresponding author. Tel.: +1 2027847182.
[1] Present address: Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30030, USA.

1755-4365 © 2014 The Authors. Published by Elsevier B.V. Open access under the CC BY-NC-SA license.
http://dx.doi.org/10.1016/j.epidem.2013.12.002

recovery times of cases in small outbreaks are correlated (Rida, 1991), which makes estimation difficult when using data from a single outbreak. An obvious solution to reducing the inaccuracy caused by within-outbreak correlations in data is to base estimates on data from multiple outbreaks.

Methods for estimating parameters from multiple outbreaks have been described before, but often have been developed for smaller data sets and computing resources than what are now available. For example, the previous approaches of Becker (1979) and Becker (1991) assumed only the observation of the final state of each outbreak was available, used moments estimators, did not formulate general a regression model to allow for variation in parameters between outbreaks, and may be implemented with pencil and paper. Our norovirus outbreak data set includes the full observation of a large number of outbreaks and a number of covariates that are likely to affect parameters. We thus here employ a different method that operates on the full observation of outbreaks, uses maximum-likelihood estimators, models the effect of covariates on outbreak parameters within a general regression framework, and exploits modern computing power to find estimates and their confidence intervals.

We propose a general approach to fitting data from many small outbreaks. Using simulated data, we assessed the performance of the proposed method as a function of the number of outbreaks in the data, the rounding of measurements to regular intervals of observation, the number of missing observations, and the imputation of missing observations. When the number of outbreaks was large, we found the performance to be satisfactory for data sets with realistic levels of all of these challenging features. Fitting our model to data from a large number of real norovirus outbreaks in health-care facilities, we found a distinct increase in transmission and initial growth rates in long-term-care facilities relative to hospitals. We examined the fit of the model and found the most noticeable defect to be lower-than-observed prediction of the initial growth of the outbreaks. However, the predicted dynamics became more accurate over time such that predictions never deviated widely from observations.

Methods

We developed the methods described in this section to fit a model of the outbreak dynamics of norovirus based on data from a large survey of gastroenteritis in health-care facilities in the former County of Avon, England. In this study, the events of symptom onset and recovery were recorded on a daily basis for cases of gastroenteritis in both care staff and patients in 15 hospitals and 135 long-term-care facilities over a year-long period in 2002–2003, and these events were classified into a total of 271 separate outbreaks (Lopman et al., 2004b). These outbreaks were for the most part small; the range in total cases spans from 2 to 90 cases and the median is 13 cases.

We begin by presenting our estimation methods. With the method defined, we then describe assumptions and imputation procedures used to prepare our data for application of the method. To complete the model specification for our application, we next describe the variables of the data chosen to be predictors of how parameters vary among outbreaks. Finally, we provide details about methods of simulation, calculation of confidence intervals, and choice of software.

Model

Although our aim is to introduce a general approach, we aim to do so by way of example. Thus we describe our methods in terms of a specific model choice made for the norovirus data. However, we do provide references to relevant results in the regression literature to indicate the full scope of this approach.

The states and transition rules for the model we adopt for individual outbreaks are as follows. The population consists of a fixed number of people of one or more types. The term type here identifies people by the rules governing their movement between different states with respect to norovirus infection. At the beginning of an outbreak, there is some positive number of people in an *exposed*, or *latent*, state for at least one of the types. This state represents people who have been exposed to an infection source and have a latent infection but are not contagious. They move to an *infective* state after an incubation period of fixed duration. The infective state represents contagious people, and for simplicity we assume that all contagious people are symptomatic. A *susceptible* state represents people who are susceptible to infection. Thus each susceptible of type i moves to the latent state at the first point of a Poisson process with rate $\beta_i Y(t)$, where β_i is the *transmission rate* for type-i susceptibles and $Y(t)$ is the number of infectives at time t. All infective types have the same level of contagiousness and have gamma-distributed symptomatic periods with the same dispersion parameter, but the mean symptomatic period may differ between types. Further, types that represent care staff are moved into an *infective-but-removed* state when the time they have spent in the infective state exceeds a threshold of fixed duration. This transition rule represents the effect of infection-control policies that prevent staff from working when contagious. At the end of their symptomatic periods, infective and infective-but-removed people are moved into a *recovered* state. The recovered state represents individuals that gain immunity over the course of the outbreak. The outbreak ends when the number of infected people reaches zero.

In summary, our outbreak model is the widely studied susceptible-exposed-infective-recovered (SEIR) model with four customizations for our application. First, we allow people to vary in susceptibility and expected duration of infectiousness. Second, we do not make our transmission rate depend on the total number of people in the population. This departure prevents the need for the total number of people to be estimated, and it is appropriate in small populations when an infective person may be able to infect every susceptible person in the population with approximately the same probability. For example, Forrester and Pettitt (2005) did not find that inclusion of the total population size significantly improved the fit of a model of methicillin-resistant *Staphylococcus aureus* (MRSA) outbreaks within an intensive-care unit. Third, we do not assume that latent periods and infectious periods are exponentially distributed. Our approach is more realistic because it allows the probability of a person leaving a latent or infectious state to depend on how long she has been in that state. Fourth, we shunt some of the infectives into an infective-but-removed state to represent the isolation of contagious staff from the population.

As indicated in our outbreak model description, the rate at which a susceptible acquires infection from an infective may vary among members of a population, and we use the word type in a general sense to refer to subsets of the population that are assumed to be the same with respect to such variation. With multiple-outbreak data, we further define types as unique to individual outbreaks. In other words, we make no general assumption that people in different outbreaks may be modeled with the same parameters. We shall later choose a particular linear model that controls the extent to which parameters may vary among types, but many other choices for such models are possible within this framework. Types thus represent the fundamental unit of variation in this framework, and the likelihood function naturally breaks apart into factors for each type.

For each type, the recovery-time and transmission-time parts of the likelihoods further factor apart into common density functions. The simplicity of these functions belies an involved construction,

Download English Version:

https://daneshyari.com/en/article/2813584

Download Persian Version:

https://daneshyari.com/article/2813584

Daneshyari.com