Vaccine 35 (2017) 5835–5841

Contents lists available at ScienceDirect

Vaccine

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

An ensemble approach to predicting the impact of vaccination on rotavirus disease in Niger

Jaewoo Park [a], Joshua Goldstein [b], Murali Haran [a,*], Matthew Ferrari [c]

[a] Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA
[b] Social and Data Analytics Laboratory, 900 N Glebe Rd, Virginia Tech, Arlington, VA 22203, USA
[c] Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 4 May 2017
Received in revised form 16 August 2017
Accepted 5 September 2017

Keywords:
Rotavirus
Vaccine
Niger
Susceptible-Infectious-Recovered models
Bayesian model averaging

ABSTRACT

Recently developed vaccines provide a new way of controlling rotavirus in sub-Saharan Africa. Models for the transmission dynamics of rotavirus are critical both for estimating current burden from imperfect surveillance and for assessing potential effects of vaccine intervention strategies. We examine rotavirus infection in the Maradi area in southern Niger using hospital surveillance data provided by Epicentre collected over two years. Additionally, a cluster survey of households in the region allows us to estimate the proportion of children with diarrhea who consulted at a health structure. Model fit and future projections are necessarily particular to a given model; thus, where there are competing models for the underlying epidemiology an ensemble approach can account for that uncertainty. We compare our results across several variants of Susceptible-Infectious-Recovered (SIR) compartmental models to quantify the impact of modeling assumptions on our estimates. Model-specific parameters are estimated by Bayesian inference using Markov chain Monte Carlo. We then use Bayesian model averaging to generate ensemble estimates of the current dynamics, including estimates of R_0, the burden of infection in the region, as well as the impact of vaccination on both the short-term dynamics and the long-term reduction of rotavirus incidence under varying levels of coverage. The ensemble of models predicts that the current burden of severe rotavirus disease is 2.6–3.7% of the population each year and that a 2-dose vaccine schedule achieving 70% coverage could reduce burden by 39–42%.

1. Introduction

Diarrheal disease is the second leading cause of death around the world for children under 5 years of age [4]. Though there are many infectious causes of diarrheal disease in children, rotavirus is the leading cause of gastroenteritis [6,39,37]. In many countries, better sanitation, hygiene and access to care have reduced the burden of diarrhea [9,20]. Despite this trend, the proportion of diarrheal hospitalizations attributable to rotavirus increased [28]. The recent development of new prophylactic vaccines for rotavirus is a promising advance in the prevention of diarrheal disease and the reduction of overall childhood mortality [29,23].

Observation of rotavirus dynamics and estimation of the burden of rotavirus disease is limited both by non-specific surveillance and under-reporting. The dynamics of rotavirus transmission must often be inferred from non-specific temporal surveillance of diarrheal disease that includes multiple causes. This is analogous to the dynamics of specific influenza strains, which are commonly inferred from non-specific time series surveillance of influenza-like illness (ILI) that includes infection by multiple influenza strains (influenza A and B), as well as additional viral infections, for example parainfluenza, coronavirus, rhinovirus [33,8]. In sub-Saharan Africa, the cause of diarrheal disease is often unknown due to a lack of diagnostic capacity [26]. Even when the cause of disease is known, an unknown fraction of cases will occur in the community and never be recorded by the health system, leading to a potentially significant level of under-reporting. Dynamic models in general, and so-called state-space models in particular, have been an important tool in the assessment of disease burden from non-specific or imperfect surveillance [17,14,5]. We estimate the burden of rotavirus in the Maradi region of Niger by synthesizing two sources of data. We use hospital surveillance data collected by Epicentre for the incident cases over time, including lab-confirmation to assess the likelihood that a case of severe diarrhea is caused by rotavirus. In addition, we use a cluster survey of households conducted to estimate the proportion of diarrheal

* Corresponding author.
E-mail addresses: jzp191@psu.edu (J. Park), joshg22@vbi.vt.edu (J. Goldstein), muh10@psu.edu (M. Haran), mjf283@psu.edu (M. Ferrari).

http://dx.doi.org/10.1016/j.vaccine.2017.09.020
0264-410X/

disease cases in the region seeking care. The latter is used to help estimate the reporting rate.

State-space models rely on the temporal correlation in a dynamic model to make the unobservable true state of the system, that is, the incidence of the pathogen of interest, estimable from noisy or imperfectly sampled data [19]. Thus, the inference about disease burden is conditional on the structure of the underlying dynamic model. For pathogens with well characterized epidemiology, such as measles and influenza, the application of state-space models to infer disease burden and transmission dynamics has become common [17,7,5,35]. The dynamics of rotavirus, which itself comprises multiple strains that result in varying levels of cross-protective immunity to other strains, has been variously described by a suite of different models [30]. Therefore, inference about rotavirus burden is limited both by imperfect surveillance of rotavirus infection and uncertainty about the underlying transmission dynamics. Rather than condition our analysis on any one model, we fit the observed time series to a suite of 5 different model structures and assumptions to account for uncertainty in model parameters as well as the dynamics represented in the models themselves.

While the development of several novel rotavirus vaccines is a promising advance for the control of diarrheal disease in children, the potential impact of the introduction of these vaccines at the population-scale is uncertain. The predicted impact of vaccine introduction may depend both on the efficacy of the vaccine and model structure; for example [30] proposed alternative models for boosting of immunity following sequential exposure to rotavirus. Bayesian model averaging (BMA) [2,16] allows for the integration of predictions of multiple models, weighted by their posterior support, to generate a single ensemble estimate that accounts for uncertainty in model selection. Here, via BMA, we use the ensemble of fitted models to predict the short-term and long-term impact of vaccination on rotavirus incidence. We then estimate the predicted impact using the vaccine efficacy from two different studies. Our ensemble approach predicts that the current burden of severe rotavirus disease is 2.6–3.7% of the population each year and that a 2-dose vaccine schedule achieving 70% coverage could reduce burden by 39–42%.

2. Material and methods

We use data from two sources: a time series of clinic admissions for diarrheal disease and a community based survey of health-seeking behavior. Clinic surveillance covers a collection of health centers and district hospitals from four districts in the Maradi region of Niger including Aguie, Guidan Roumdji, Madarounfa, and the city of Maradi. A total of 9590 cases of diarrhea in children under 5 were recorded from December 23, 2009 to March 31, 2012 (118 weeks). For each patient age in months, village of origin, date of consultation were recorded. Also noted were potential symptoms including temperature, duration of diarrhea before consultation, presence of blood in the stool, presence and duration of vomiting, and level of dehydration. In each case a rapid test was administered for detecting rotavirus. Using the rapid test, 2921 cases tested positive for rotavirus. A subset of 378 cases testing positive for rotavirus were also genotyped. While 32 separate strains were identified, more than two thirds of positive cases were of strains G12P[8] or G2P[4].

The distributed nature of Niger's healthcare system is a challenge for surveillance. Roughly a third of all health centers in these districts were included. Notably absent were the many local health posts staffed by community health workers. To estimate both the fraction of cases seeking care at a health center, and the fraction seeking any level of care, a second source of data is needed. We use a community survey [27] of children approximately under 5 years old to get estimates of these reporting rates.

A total of 2940 children under 5 were selected for inclusion in the cluster survey from households across the four districts. Clusters were allotted according to the population of each village from census data. Sampling weights accounted for household composition and the relative populations of the districts. Among those surveyed, 1099 caregivers reported at least one episode of diarrhea during the recall period of 27 days. Respondents reported whether they sought care at a health structure. We use the reporting rate of severe diarrhea, which is defined as the presence of acute watery diarrhea and the presence of two or more of the signs of loss of consciousness, sunken eyes, and an incapacity to drink or drinking very little.

From the cluster survey we determine that an estimated total of 42.9% of caregivers who reported severe diarrhea consulted at a health center ($95\%\text{CI} : (33.1\%, 52.7\%)$). The rest either sought care at a district hospital, local health post or did not seek care at a formal health structure. This estimate is used as a proxy for the reporting rate of rotavirus. More specifically, this information is used to construct an informative prior for our Bayesian approach (as described in the supplementary material).

2.1. Model overview

We consider a range of dynamic models for rotavirus transmission. Information linking individual-level data on the course of infection to the between-person transmission of rotavirus is lacking, leading to variation in the structure of mathematical models for rotavirus [30]. Using a range of different models allows us to account for the uncertainty in estimation due to model choice. The five models we consider are SIR-like compartmental models of transmission, building upon the models in [30]. While the structure of our model is the same as [30], there are two distinctions: (1) our focus is on age-structured modeling for children under the age of five instead of across all age groups; (2) we use a Bayesian inferential approach (as opposed to the maximum likelihood approach in [30]). The latter difference is crucial because we use Bayesian methodology to obtain posterior model probabilities for each of five potential model structures given the observations. This not only enables a probabilistic comparison of the different models but also allows for Bayesian model averaging, thereby providing an ensemble-based projection of rotavirus burden as well as the impact of vaccination. We incorporate age into the model with separate compartments for ages from 0–1 month, 2–3 months, 4–5 months, 6–11 months, 12–23 months, and 24–59 months. Fixed parameters including infection period, immunity period, and exposed period in the SIR models are obtained from Table 2 in previous work [30]; these estimates are from data from England and Wales.

Here we very briefly outline the main features of five models, Models A through E, based on the SIR framework. Details of the model and inferential procedure are described in the supplementary material. Model A tracks severe and mild rotavirus separately. Severe infections disproportionately contribute to the force of infection (e.g. because of higher rates of shedding [30]). Unlike Model A, Models B-E assume successive infections and immunity are obtained through repeated infections. Subsequent infections will have a reduced susceptibility to infection and level of infectiousness. Model C allows for an incubation period of infections as well. In Model D there is no temporary immunity during successive infections and immunity is granted after all repeated infections. Model E assumes that full immunity can be obtained during successive infections. In all models, we assume that the transmission varies as a cosine function with a period of 1 year; the mean and amplitude of that function are parameters to be