Journal of Theoretical Biology 420 (2017) 68–81

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

# Model distinguishability and inference robustness in mechanisms of cholera transmission and loss of immunity

Elizabeth C. Lee[a], Michael R. Kelly Jr[b,c], Brad M. Ochocki[d], Segun M. Akinwumi[e], Karen E.S. Hamre[f], Joseph H. Tien[c], Marisa C. Eisenberg[g,*]

[a] Department of Biology, Georgetown University, 37th and O Streets, NW, Washington, DC 20057, United States
[b] Division of Natural Sciences and Mathematics, Transylvania University, 300 North Broadway, Lexington, KY 40508, United States
[c] Department of Mathematics, The Ohio State University, 231 West 18th Ave, Columbus, OH 43210, United States
[d] Department of BioSciences, Program in Ecology and Evolutionary Biology, Rice University, 6100 Main Street, MS-170, Houston, TX 77005-1892, United States
[e] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada T6G 2G1
[f] Division of Global Pediatrics and Division of Epidemiology and Community Health, University of Minnesota, 717 Delaware Street SE, 3rd Floor, Minneapolis, MN 55414, United States
[g] Departments of Epidemiology and Mathematics, University of Michigan, Ann Arbor, 1415 Washington Heights, Ann Arbor 48109, United States

## ARTICLE INFO

*Keywords:*
Model misspecification
Parameter estimation
Model structure
Comparative modeling

## ABSTRACT

Mathematical models of cholera and waterborne disease vary widely in their structures, in terms of transmission pathways, loss of immunity, and a range of other features. These differences can affect model dynamics, with different models potentially yielding different predictions and parameter estimates from the same data. Given the increasing use of mathematical models to inform public health decision-making, it is important to assess model distinguishability (whether models can be distinguished based on fit to data) and inference robustness (whether inferences from the model are robust to realistic variations in model structure).

In this paper, we examined the effects of uncertainty in model structure in the context of epidemic cholera, testing a range of models with differences in transmission and loss of immunity structure, based on known features of cholera epidemiology. We fit these models to simulated epidemic and long-term data, as well as data from the 2006 Angola epidemic. We evaluated model distinguishability based on fit to data, and whether the parameter values, model behavior, and forecasting ability can accurately be inferred from incidence data.

In general, all models were able to successfully fit to all data sets, both real and simulated, regardless of whether the model generating the simulated data matched the fitted model. However, in the long-term data, the best model fits were achieved when the loss of immunity structures matched those of the model that simulated the data. Two parameters, one representing person-to-person transmission and the other representing the reporting rate, were accurately estimated across all models, while the remaining parameters showed broad variation across the different models and data sets. The basic reproduction number ($\mathcal{R}_0$) was often poorly estimated even using the correct model, due to practical unidentifiability issues in the waterborne transmission pathway which were consistent across all models. Forecasting efforts using noisy data were not successful early in the outbreaks, but once the epidemic peak had been achieved, most models were able to capture the downward incidence trajectory with similar accuracy. Forecasting from noise-free data was generally successful for all outbreak stages using any model.

Our results suggest that we are unlikely to be able to infer mechanistic details from epidemic case data alone, underscoring the need for broader data collection, such as immunity/serology status, pathogen dose response curves, and environmental pathogen data. Nonetheless, with sufficient data, conclusions from forecasting and some parameter estimates were robust to variations in the model structure, and comparative modeling can help to determine how realistic variations in model structure may affect the conclusions drawn from models and data.

* Corresponding author.
*E-mail addresses:* ecl48@georgetown.edu (E.C. Lee), mikelly@transy.edu (M.R. Kelly), brad.ochocki@rice.edu (B.M. Ochocki), segunmic@ualberta.ca (S.M. Akinwumi), hamr0091@umn.edu (K.E.S. Hamre), jtien@math.ohio-state.edu (J.H. Tien), marisae@umich.edu (M.C. Eisenberg).

http://dx.doi.org/10.1016/j.jtbi.2017.01.032
Received 22 May 2016; Received in revised form 16 January 2017; Accepted 19 January 2017
Available online 24 January 2017

## 1. Introduction

Cholera is a waterborne disease caused by the bacterium *Vibrio cholerae*, which manifests as severe diarrhea and vomiting leading to dehydration. Left untreated, cholera can be up to 50% fatal, but rehydration treatment can greatly reduce case fatality rates to as low as 1% (Sack et al., 2004; WHO, 2011). Worldwide, cholera causes three to five million cases and over 100,000 deaths per year (World Health Organization, 2010). Numerous mathematical models of cholera transmission have been proposed to investigate factors that impact the dynamics and transmission of waterborne diseases (Codeco, 2001; Hartley et al., 2006; King et al., 2008; Joh et al., 2009; Tien and Earn, 2010; Mwasa and Tchuenche, 2011; Shuai et al., 2012; Tian and Wang, 2011; Andrews and Basu, 2011; Sanches et al., 2011), and the ongoing cholera epidemic in Haiti has spurred additional interest in the subject (Tuite et al., 2011; Chao et al., 2011; Abrams et al., 2012; Andrews and Basu, 2011; Bertuzzo et al., 2011; Eisenberg et al., 2013a).

Due to the range of indirect transmission pathways and timescales, which may be represented by environmental water sources, household water containers, foodborne transmission, and more (Swerdlow et al., 1992; Eisenberg et al., 2013b), commonly used mathematical models for cholera vary widely in their level of detail, spatial scale, and model structure. Some models use a single term to represent a composite set of transmission mechanisms, while others include multiple timescales of transmission (Codeco, 2001, Tien and Earn, 2010). In addition to standard mass-action transmission (Tien and Earn, 2010; Tuite et al., 2011), many models use nonlinear transmission functions for waterborne transmission (Codeco, 2001, Mukandavire et al., 2011) to reflect the dose response for cholera in the water. Models may also include an asymptomatic transmission pathway, a hyperinfectious state for the bacteria immediately after shedding, and other ecological and environmental factors in the environmental reservoir, such as effects of vibriophages, plankton, weather, and climate (Tien and Earn, 2010; King et al., 2008; Codeco, 2001; Hartley et al., 2006; Shuai et al., 2012; Pascual et al., 2006; Koelle et al., 2005; Eisenberg et al., 2013b;2013a; Miller-Neilan et al., 2010; Rinaldo et al., 2012; Akman and Schaefer, 2015).

In addition to the variation in transmission mechanisms, loss of immunity to cholera is poorly understood and therefore, modeled with many different assumptions. Estimates of the length of immunity in the literature range widely from several months to three to ten years (Levine et al., 1981; King et al., 2008; Koelle and Pascual, 2004). Immunity to cholera is of particular interest given the recent and ongoing oral cholera vaccine campaigns worldwide, including in Haiti, Bangladesh, and Thailand (Tohme et al., 2015; O'Leary and Mulholland, 2015; Deen et al., 2016; Phares et al., 2016), which raise additional questions of how vaccine-derived immunity compares to immunity derived from infection.

As modeling gains prevalence among policy makers in public health (Abrams et al., 2012; Moyer, 2014; Auchincloss and Roux, 2008; Grad et al., 2012; Lofgren et al., 2014; Lipsitch et al., 2011), comparative or ensemble modeling approaches have been increasingly viewed as a way to ensure that the results of parameter estimation, forecasting efforts, and the evaluation of intervention strategies are conserved across the range of realistic model structures (Koopman, 2004; Meza et al., 2014). Two related concepts are useful to consider in these efforts—*model distinguishability* addresses whether candidate models can be distinguished by their fits to empirical data (Walter et al., 1984), and *inference robustness assessment* examine whether conclusions drawn from a particular model are robust to realistic variations in the assumptions and model structure (Koopman, 2004). Both of these concepts are important to evaluate in the model-building process when model results are used as the basis for decision making. While parameter uncertainty has been highlighted for cholera outbreaks both in general (Fung, 2014; Akman and Schaefer, 2015) and in Haiti-specific contexts (Fung, 2014; Rinaldo et al., 2012), model distinguishability is a higher order examination of uncertainty at the model structure level. In previous work (Eisenberg et al., 2013b), we examined identifiability of a very simple but commonly used environmental transmission model of cholera, examining structural (theoretical) and practical identifiability in the context of distinguishing between direct (non-environmentally driven) and indirect (environmentally-driven) cholera transmission. However, the model considered there did not include a range of real world features known in cholera epidemiology (e.g. loss of immunity, dose response, asymptomatic cases) (Fung, 2014), and relatedly it did not consider issues of model distinguishability (beyond considering direct and indirect transmission), which are highly relevant given the real-world complexity and range of models in the literature.

In this paper, we examine the effects of uncertainty in model structure on cholera disease dynamics and inference by considering five models with different transmission and loss of immunity mechanisms. The models under consideration share a common base, the SIWR model of Tien and Earn (2010). The SIWR model is an extension of the classic Susceptible-Infected-Recovered (SIR) (Kermack and McKendrick, 1927) with an added compartment for pathogen concentration in an aquatic reservoir (W) (Tien and Earn, 2010). In addition to the person-to-person and water transmission pathways of the base SIWR model, we evaluate two additional transmission-related features: a Michaelis-Menten dose response term for waterborne transmission and an asymptomatic infection and transmission pathway (Dunworth, 2011; Codeco, 2001; King et al., 2008; Miller-Neilan et al., 2010). We also consider three loss of immunity features not included in the SIWR model: exponential loss of immunity, stage-progression gamma-distributed loss of immunity, and a novel model that features progressively increasing susceptibility after recovery from infection.

Using these five deterministic SIWR-based models, we first simulate data from each model with different types of added noise. In the frame of model distinguishability and inference robustness assessment, we estimate model parameters and fit model data for all five models to each simulated dataset and an empirical dataset from the 2006 cholera epidemic in Angola. We determine how well each model recaptures the underlying parameter and $\mathcal{R}_0$ values, as well as how each model fits to simulated data generated from different models. Finally, we forecast trajectories using parameters estimated from truncated simulated incidence data and compare model forecasts with that of the true simulated data.

## 2. Methods

### 2.1. Model descriptions

Here we introduce the five models of epidemic cholera disease dynamics that were used in our analyses. We base these models on the SIWR model of Tien and Earn (2010), which includes two timescales of transmission: a fast, direct route (represented by $\beta_I$) which incorporates direct person-person transmission, foodborne and household transmission, and other non-environmental transmission pathways; and a slow, indirect route ($\beta_W$) which represents long-term transmission mediated by bacteria in the water. The scaled non-dimensional model equations are given by Tien and Earn (2010):

$$\begin{aligned} s' &= \mu - \beta_I si - \beta_W sw - \mu s \\ i' &= \beta_W sw + \beta_I si - \gamma i - \mu i \\ w' &= \xi(i - w) \\ r' &= \gamma i - \mu r \end{aligned} \tag{1}$$

where $s$, $i$, and $r$ represent the fractions of the population that are susceptible, infectious, and recovered, respectively. The $w$ variable is proportional to the concentration of *V. cholerae* in the environment, and $\xi$ represents the decay rate of the pathogen in the water (noting that the nondimensionalization in Tien and Earn (2010) rescales the