Challenges to quantitative modelling of cholera disease transmission

Yonatan Grad - Instructor in Medicine, Division of Infectious Diseases, Brigham and Women’s Hospital, Harvard Medical School, and Research Fellow, Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health

Joel C. Miller - Assistant Professor of Mathematics and Biology, Penn State University

Marc Lipsitch - Professor of Epidemiology and Director, Center for Communicable Disease Dynamics, Harvard School of Public Health

Cholera outbreaks

In October 2010, a cholera epidemic erupted in Haiti, spreading rapidly throughout the country. Conditions were favourable for cholera spread due to poor access to clean water and sanitation caused by Haiti’s poverty compounded by severely damaged infrastructure due to the earthquake that took place 10 months before. Estimates of cholera cases in Haiti now stand at over 500,000 and of cholera deaths at over 7,000. Similarly, in late 2008, starting at a time of economic crisis, an outbreak of cholera spread throughout Zimbabwe, resulting in over 100,000 cases and over 5,000 deaths.

During these outbreaks, response efforts centred on the mainstays of treatment – antibiotics and oral and intravenous rehydration therapy – and on improving access to clean water and sanitation. Though cholera vaccines exist, they were not used for a number of reasons, one of them being that vaccines remain untested in epidemic settings and are of uncertain efficacy in populations that lack prior exposure and immunity, like in Haiti. In addition, there was concern that efforts to administer an unproven vaccine requiring two doses two weeks apart may divert resources from proven treatments.

Cholera transmission models

In this context, several groups proposed mathematical models of cholera transmission (1-5). Their goals were to predict the impact of vaccination and other interventions, and thereby to provide guidance to policy-makers on how to efficiently allocate resources during cholera epidemics. By fitting the transmission model to data from the early stages of a growing epidemic, models can estimate key parameters, like $R_0$ (the basic reproductive number, which is the number of secondary infections that arise from a typical infected individual immediately after introduction to a completely susceptible population). Running the model forward in time allows modellers to make predictions about how an epidemic will progress and about the impact of various control strategies.

Benefits from an intervention are a combination of direct effects on those receiving the intervention and indirect effects on those with reduced exposure because others received the intervention. In the context of vaccination, producing immunity through vaccination of only a part of the population may stop an epidemic because chains of transmission are broken – a concept referred to as “herd protection”. The fraction of the population one would need to vaccinate effectively to stop an epidemic, also known as the critical vaccination threshold, is $1/1/R_0$.

All models are mathematical formulations of assumptions that are thought to underlie a given process. Thus confidence in a model’s conclusions depends on confidence in the model’s assumptions. We describe briefly below and in greater detail in our article (6) how these assumptions pose challenges to making quantitative predictions, highlighting the uncertainties in model structure and parameter estimates and the potential impact of spatial heterogeneity of cholera incidence.

A mathematical model should capture the main dynamics of disease transmission and should be based on parameters from data when available. In the classic mass-action model Susceptible-Infected-Recovered (or SIR) compartmental model, the population is divided into those three classes (7). To describe how those populations change over time, equations relate the compartments through a set of parameters including population size, the rate of contact between infectious and susceptible individuals that leads to susceptible
individuals becoming infected and infectious, and the rate of recovery from infection.

In the case of cholera, transmission depends on infection by ingesting cholera bacteria in contaminated food or water and then contamination of food and water by infectious bacteria excreted by the infected individual. In 2001, Codeço (8) proposed a cholera model that accounts for waterborne transmission by introducing an additional compartment for cholera in the water supply — termed an SIWR model. This model attempts to capture the role of contaminated water: individuals become infected by drinking cholera-containing water and infected individuals then propagate the infection by shedding cholera bacteria via their stool into the water supply. Once in the water supply, bacteria start to die off. Each of these rates – rate of water supply contamination, rate of contact with the contaminated water, and rates of loss of the pathogen in the water supply – requires additional parameters.

**Challenges in making accurate predictions**

Many of these variables remain unknown or poorly described, leading to considerable uncertainty in the appropriate parameter values to use when trying to apply this model to a specific location like Haiti or Zimbabwe. The rate of contamination of the water supply, for example, must reflect the size of the water reservoir, the amount of the person’s infected stool that reaches the reservoir, and the concentration of cholera bacteria in the stool. This is essentially a summary of the severity of infection, local sewage infrastructure, and the kind and size of the water supply, all of which are empirically unknown and likely to vary.

As another example, take the parameter that describes how long cholera bacteria last in the water supply. This is a difficult parameter to characterize, and it varies in response to numerous factors, including water salinity, temperature, exposure to sunlight, and the amount of cholera phage – viruses that kill cholera – in the local environment (9-12). The local concentration of cholera bacteria may also be subject to the type of water supply itself: cholera bacteria in a river may be swept away by the current, in contrast to cholera in a well. And yet this parameter has significant implications for the predictions generated by the model. Using the SIWR model described above, assuming an average lifespan of 3 days for cholera bacteria in the water supply and fixing the other parameters, yields an $R_0$ of 1.95. Changing the average lifespan to 30 days while holding all the other parameters equal results in an $R_0$ of 6. This uncertainty in the lifespan of cholera then translates to significant uncertainty in population-level effects of the interventions, where the herd protection benefit of vaccination could extend from complete protection of the unvaccinated population to almost no protection.

An additional challenge to making accurate predictions from mathematical models based on data aggregated over cities or provinces comes from the heterogeneous behaviour of the disease and its transmission across locations. The spatial heterogeneity of cholera is well known, dating to John Snow’s classic investigation into why one area of London had a much higher rate of cholera than adjacent neighbourhoods. Similar variations across city neighbourhoods were observed in Harare during the 2008-9 outbreak (13), and these trends extend to larger geographic scales, with cumulative attack rates in Haiti as high as near 30% in some districts and as low as less than 1% in other districts (14). This extent of spatial heterogeneity can confound efforts to estimate parameter values from aggregated data, as model fitting to aggregated data assumes that the parameters can be applied homogeneously. While fitting a model to aggregated data will generate a single $R_0$, this value is likely to be an underestimate for any given community, so conclusions about what would be an adequate level of intervention might be sufficient for control in one constituent area but inadequate in other areas. Since estimates for the critical vaccination threshold depend on $R_0$, as described above, aggregation of spatially heterogeneous data may lead to misleading conclusions about the potential impact of vaccination.

Models can be considered a hypothesis-generating tool, in addition to predictors of the effects of interventions on transmission. The sources of uncertainty described above suggest new avenues for research, from types of data that should be collected to address parameter uncertainty to the sensitivity analyses that should be conducted to address parameter variability. For example, higher spatial resolution mapping of cases will help address the problem of spatial heterogeneity, and detailed characterization of water resources and their use will improve parameterization and understanding of how critical parameters vary by location.

Now that the World Health Organization has been given a mandate to generate an oral cholera vaccine stockpile for use in epidemic settings, modelling of cholera transmission takes on a new importance. Modelling may be helpful with key questions: Under what conditions should one consider vaccination of a population? How can one maximize the potential impact of the vaccine, using a given amount of vaccine to benefit the largest number of people, including those who are not directly impacted by vaccination and those indirectly by herd protection? Ongoing efforts to employ vaccines in epidemic settings provide valuable opportunities for collecting key data to improve the utility of cholera models in guiding policy decisions.

**Yonatan Grad**, Instructor in Medicine, Division of Infectious Diseases, Brigham and Women’s Hospital, Harvard Medical School, and Research Fellow, Center
for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health. Research interests: development and application of mathematical modeling and genomic epidemiology tools to study the emergence and spread of pathogens. ygrad@hsph.harvard.edu

Joel C. Miller, Assistant Professor of Mathematics and Biology, Penn State University. Research interests: the development and analysis of mathematical models of infectious disease spread. joel.c.miller.research@gmail.com

Marc Lipsitch, Professor of Epidemiology and Director, Center for Communicable Disease Dynamics, Harvard School of Public Health. Research interests: Antigenic variability and immunity, Antimicrobial resistance, Influenza, Pathogen genomics, Streptococcus pneumoniae, Transmission-dynamic modeling. mlipsitch@hsph.harvard.edu

References: