Building Capacity for Meaningful Epidemiological Modelling

Steve Bellan – PhD candidate Department of Environmental Science, Policy & Management, University of California, Berkeley, USA.

Mathematical modelling of infectious diseases

Mathematical modelling is valuable in public health because it provides a way to evaluate strategies for controlling disease before actually trying such strategies in the field. It also allows researchers to explore hypotheses about how a disease is transmitted, by allowing them to determine how the disease would behave under hypothetical scenarios. More generally, mathematical modelling provides a rigorous way to describe complicated systems, where many factors interact interdependently. The transmission of an infectious disease involves interactions between people. For a given disease, the rate at which infections arise depends on past interactions between infected and susceptible people. Thus, the rate at which people become infected depends on the history of the infection in the population. This motivates the utility of mathematical modelling in infectious disease epidemiology. Dynamic mathematical modelling caters for the analysis of systems in which rates of change depend on the quantities that are themselves changing.

Using calculus - the type of math used to describe rates of change - equations can be written that describe how interdependent populations change together. For instance, the classical SIR (Susceptible, Infected, Recovered) model, developed by researchers in the 1920s, is a basic epidemiological model in which three equations describe how the number of susceptible, infected, and recovered people in a population change over time (1). The trajectory of an epidemic is determined once the model's parameters are specified. In this case, the parameters are simply numbers describing the prevalence of infection at time zero (the beginning of the model run), the contact rate between people, the probability of infection given contact between a susceptible and an infected person, and the duration of the infectiousness.

Depending on the question to be addressed, the SIR model may be elaborated or entirely different models may be formulated to allow for other phenomena such as random chance, individual variability, and many other complexities that may be of relevance. Models can be developed for very specific populations, or to assess generic principles of disease transmission and intervention in a generalized population. By varying the values of model parameters, it is possible to ask questions such as: "How will the number of people getting sick during an influenza outbreak in Cape Town be affected if we give half of all sick people a drug that reduces their infectiousness by 50%?"; "How do the results change if we assume some individuals are much more infectious than others to begin with?"

Mathematical modelling of infectious diseases should be done in close collaboration with epidemiologists and public health professionals. Useful epidemiologic models are those aimed at answering important questions in public health and informed by real-world knowledge of the diseases if interest. As such, models must be sufficiently realistic and complex to represent the factors relevant to the question being addressed, while otherwise remaining as simple as possible. Fitting models to epidemiologic data, when available, provides one important way to ground models in reality. Fitting models to data, however, requires a firm understanding of how data is generated and analysed in epidemiologic research.

Clinic on Meaningful Modelling of Biological Data

In May 2009, 41 researchers and students from across Africa and North America met for 9 days at the African Institute for Mathematical Sciences (AIMS) in Muizenberg, South Africa for the first annual clinic on the Meaningful Modelling of Biological Data. The clinic brought together mathematicians, statisticians, ecologists and epidemiologists, to engage with meaningful questions about infectious disease dynamics by integrating mathematical models with epidemiological data. The emphasis of the clinic was to focus on a "bottom up" rather than a "top down" approach. That is to say, primary attention was focused on the data, allowing the data to inform the construction of the simplest models appropriate, rather than on the development of complex mathematical models unrelated to data. The success of this first, experimental, clinic suggested the need for a follow-up and, accordingly, the second clinic on the Meaningful Modelling of Epidemiological Data was held at AIMS on May 24 – June 4, 2010.

The clinic was developed out of the observation that many mathematicians, particularly in Africa, are highly motivated to apply their mathematical modelling skills to pressing public health problems but

often have trouble bridging the gap between theory and real-world applications. In particular, they lack the skills necessary to analyse data and use it to develop and test their models. The primary focus of these clinics is accordingly to illustrate techniques for bridging the gap between models and data, using a series of interactive lectures and computer tutorials, moving gradually from canned exercises to independent exploration of novel research ideas.

Focus was for students to develop their own research projects and linking theoretical work to available data sets. Lectures were linked closely to computer exercises to help participants master a data-driven modelling approach. Participants worked through a set of exercises that illustrated how to fit a simple mathematical model of HIV transmission to a time series of HIV prevalence taken from antenatal clinic surveys conducted over two decades in Uganda. They were asked to think about why the simplest possible model could not satisfactorily explain observed prevalence trends and to progressively modify the model, building up an understanding of which assumptions were necessary to produce the observed patterns. Next participants were asked to engage with new data sets not previously explored with the tools they were using, interpret differences between the initial data set they had explored and the new data. As the clinic moved forward, the problems posed became more complex, and eventually participants were presented with active research problems. In particular, participants were given HIV testing data from antenatal clinics in Harare, Zimbabwe and were asked to think creatively about how the age structure of the testing data could be used to develop estimates of incidence (i.e. the rate at which new HIV infections are produced) from data on prevalence (i.e. a measure of the proportion of HIV positive individuals amongst those tested) and about how antenatal clinic data can best be used as a surrogate for prevalence in the whole population.

One of the goals of the clinic was to teach the philosophy that motivates this type of work. With that in mind, a substantial amount of time was spent asking participants to simply look at data, describe what they saw, and discuss potential drivers of observed patterns - based not on formal mathematical models but on their intuition and real-world knowledge. In essence, the goal was to help the participants develop a skill for asking meaningful questions based on their observations and realize that once a meaningful

question was formed, a model could become a tool for formulating and testing hypotheses.

To allow participants to continue projects started at the workshop, the clinics rely entirely on open access software, including the Open Office spreadsheet program Calc (2) and the statistical programming language R (3). Additionally, most of the data used are from publicly available sources. A major goal of future clinics will be developing an interactive online research community that will facilitate long-term international collaborations between participants. By posting links to publicly available data and creating forums for their discussion and analysis, this community will also stimulate the development of new collaborations between researchers on open problems in infectious disease dynamics.

Steve Bellan, PhD candidate Department of Environmental Science, Policy & Management, University of California, Berkeley, USA. Areas of interest: the application of mathematical and statistical techniques to applied problems in infectious disease epidemiology. His current research aims at understanding the causes and consequences of anthrax outbreaks in the wildlife of Etosha National Park, Namibia. *sbellan@berkeley.edu*

Reference list
1. Smith D, Moore L. The SIR model for spread of disease. Journal of online mathematics and its applications. http://mathdl.maa.org/mathDL/4/?pa=content&sa=viewDocument&nodeId=479 Accessed May 18, 2010.
2. Open Office.org. Calc – the all-purpose spread-sheet. http://www.openoffice.org/product/calc.html Accessed May 18, 2010.
3. The R Project for Statistical Computing. http://www.r-project.org/ Accessed May 18, 2010.