BIRS Focused Workshop 07frg109, May 2007

Stochastic models of influenza dynamics

Participants:

: Sergey Kryazhimskiy (Princeton University),

: Jonathan Dushoff (McMaster University),

: David Earn (McMaster University),

: Junling Ma (University of Victoria),

: Mark Lipsitch(Harvard University),

: Joshua Plotkin (University of Pennsylvania),

: Ben Bolker (University of Florida),

: Trevor Bruen (University of Toronto)

The recent workshop at BIRS offered us a fantastic opportunity for collaboration and focused, productive research. The workshop exceeded our expectations in terms of the breadth of the academic subjects we explored, and the collaborations we established.

A subset of our group has been collaborating for several years. We have used mathematical models to study the spread and evolution of influenza viruses. The purpose of this workshop was to attempt reconciliation of our stochastic models with empirical data on influenza epidemics; to examine alternative methods of fitting model parameters; and to continue a collaboration with Marc Lipsitch from the Harvard School of Public Health. We have progress to report on all of these goals.

During our workshop at Banff, we completed revising a manuscript that uses empirical data about influenza deaths in the US over the past century to highlight a theoretical puzzle about influenza persistence after a pandemic. The most basic, longstanding mathematical model of disease transmission divides the population into three classes (Susceptibles, Infectious, and Recovered/Immune individuals) and describes flow between these classes with a system of three ordinary differential equations. Given this standard model of disease, and given the empirical influenza epidemic curve and infection rates observed in the United States in 1918, we have estimated that a very large proportion of the population was infected (and thereafter immune) to the Spanish Flu of 1918. According to these estimates, only a very small proportion of the population remained susceptible to influenza after the pandemic - too small to support the initiation of another epidemic the following season. But the empirical data indicate that another influenza epidemic did indeed occur in 1919, which raises a theoretical puzzle. Our manuscript describes this enigma and offers several hypotheses for its resolution: the virus may have evolved to such an extent in 1918 that it could re-infect individuals in 1919; or the virus could have persisted in 1919 due to heterogeneities in the host population and "pockets" of remaining susceptibles; or (perhaps most intriguing) the virus may have evolved a greater ability to spread so that it could persist in 1919, despite the small number of susceptible hosts to support it. Our manuscript does not attempt to resolve this enigma, but rather to describe how the puzzle arises from the combination of standard mathematical models and empirical data from the 1918 influenza pandemic.

The second major topic we discussed in our workshop involves the non-parametric inference of model parameters from empirical time-series data. Recent developments due to Wallinga and Teunis (2004) have allowed for direct estimates of a pathogen, reproductive number based only on a time-series of incidence counts. Unlike parametric fitting procedures, this approach is both elegant and widely applicable over a range of models. However, for pathogens like influenza we rarely have a time-series of infection events, but rather have only a time-series of death events. We spent a large amount of our time at the workshop studying how to modify the method of Wallinga & Teunis to accommodate a death time-series, instead of an incidence time-series.

One approach to this question has involved modifying the method of Wallinga & Teunis to deal with the distribution of times between one death event and another related death event. We have discovered that this approach suffers from several technical and one major conceptual difficulty. The conceptual difficulty is that a

death event on day *t* may been caused from a transmission event that eventually led to a death event on day $t' > t$. As a result, this modified version of Wallinga & Teunis's approach involved summations over both past and future events, wheras the original Wallinga & Teunis method was one-sided. This causes difficulty in many practical settings, and the sense in which the reproductive number on day t is estimated does not agree with the original Wallinga-Teunis method.

An alternative approach is to deconvolve the observed death timeseries, using knowledge of the incidence-to-death transition kernel, to impute an underlying incidence time-series which can then be analyzed by the original Wallinga-Teunis technique. In fact, we had considered this problem at a previous FRG, only to find that iteration techniques for deconvolution seemed poorly behaved. In this FRG, however, we made substantial progress in deconvoluting epidemiological data by applying the Lucy-Richardson technique, which uses an implicitly Bayesian approach that guarantees positivity at every iterate. This algorithm has not previously been applied to epidemiological data, and clearly has important epidemiological applications (recovering incidence from mortality time series, or infection from onset etc), but it has important limitations. Although the algorithm works perfectly when data are measured without noise, applications to noisy data are more complicated, involving heuristic choices to balance the smoothness of the deconvolved timeseries against the accuracy of the convolution. We are currently writing a manuscript that describes deconvolution, its relationship to Wallinga & Teunis, and its applicability to questions in disease dynamic modeling.