

*BIRS Focused Research Group*  
**Influenza Dynamics: Models and Data**  
May 28 - June 9, 2005

*Participants: Chris Bauch (University of Guelph), Jonathan Dushoff (Princeton University), David Earn (McMaster University), Junling Ma (McMaster University), Christina Mills (Harvard University), Joshua Plotkin (Harvard University)*

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 models with empirical data on influenza epidemics; and to form a collaboration with Christina Mills and Marc Lipsitch from the Harvard School of Public Health. We have progress to report on both of these goals. Perhaps most important is the strong collaboration we have formed with the Mills/Lipsitch group, resulting in two completed manuscripts already. The substance of these studies, as well as others that we initiated at Banff, are described below:

During our workshop at Banff, we completed a manuscript (MS #1) that uses empirical data from the infamous 1918 “Spanish Flu” pandemic and highlights theoretical puzzle about influenza persistence. 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 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, allowing it to persist 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.

We have also drafted a second manuscript (MS #2) that analyzes the effects of spatial aggregation of data on the estimation of critical epidemiological parameters,

such as the initial rate of disease spread, used in mathematical models. Measures of disease transmissibility are often estimated using data aggregated at a large spatial scale (e.g. city, state, country). Using 1918 influenza pandemic death data gathered at multiple spatial scales, we have shown that aggregation in the context of asynchronous epidemics of variable size tends to bias transmissibility estimates downward.

We also have begun a systematic analysis of methods used to estimate the initial rate of disease spread (a parameter called  $R_0$ ) on the basis of epidemiological data. Data available is typically either a time-series of infected individuals, a time-series of mortality events, and/or data on the probability distribution of the disease's "serial interval" – that is the duration from infection to the end of infectiousness. Aside from several standard curve-fitting methods, we developed a novel technique for estimating the rate of disease spread, based on "serial interval" data. We are planning to write a detailed, more theoretical paper (MS #3) in which we simulate standard stochastic models of disease spread, and then apply a variety of techniques to estimate the parameter  $R_0$  used in those simulations. We expect that estimates of  $R_0$  may, unfortunately, depend upon which estimation techniques are employed. We plan to investigate and present these dependencies, thereby informing the broader community of scientists and public health officials who seek to infer underlying disease parameters from epidemiological data.

Finally, in light of the three manuscripts discussed above, we are planning a fourth paper (MS #4) focused on the empirical data from the 1918 influenza pandemic in Philadelphia, which killed a staggering 12,162 people within two months. Our initial analyses of these data indicate that the epidemic time-series does not conform to the standard mathematical model of disease transmission, except during the initial few weeks of exponential growth. Instead, the Philadelphia data show a depression in the incidence rates after the first several weeks – which may suggest that behavioral changes or quarantine regulations had an important effect on curbing Philadelphia's epidemic. We intend to analyze the Philadelphia epidemic curve in detail, using methods described above, and to correlate our analysis with historical documents on the timing and extent of quarantine measures implemented in Philadelphia during the 1918 epidemic.