

Modeling the second wave of Novel soH1N1 Virus in Canada (as at 9/5/09)

Article Summary Line: Relatively stable fits to the observed Canadian soH1N1 observations provide powerful tools for public health planning, preparedness and monitoring during the pandemic.

Running Title: Modelling the second wave of soH1N1 in Canada

Keywords: Disease Outbreaks; Computer Simulation; Influenza A Virus, H1N1 Subtype;

Title: Modeling the second wave of Novel soH1N1 Virus in Canada

Authors: Paul Smetanin, David Stiff, Paul Kobak, Glen Sherman, Neil Simonsen

Author Affiliations: RiskAnalytica Research, Toronto, Ontario, Canada (P. Smetanin, D. Stiff, P. Kobak, G. Sherman); Southern Cross University, East Lismore, New South Wales, Australia (P. Smetanin), National Microbiology Laboratory, Winnipeg, Manitoba, Canada (N. Simonsen)

Abstract

Due to the capricious nature of a pandemic, the precise impact of a soH1N1 second wave remains largely a matter of speculation. This is a cause of concern for those responsible for coordinating pandemic responses. Dynamic pandemic modelling can assist resource planners that need to adapt to a pandemic as it unfolds. Since early June, a pandemic model has been periodically fit to soH1N1 Canadian data with resulting attack rates ranging between 21% and 29%, peaking generally around late December. As of September 5th, 2009 the attack rate estimate stood at 29% with an estimated case fatality rate of 0.019% and a hospitalization case rate of

0.4%. The relative stability of the current fits provides a useful resource planning tool. Changes in future fits could occur, indicating either new information or additional variables that are required to be considered, both of which also assist planners.

Introduction

The novel influenza A H1N1 virus (soH1N1) currently in circulation worldwide is the first global influenza pandemic since 1968. Given the epidemiology of pandemics vary, with disease specific and population specific uncertainty abundant, reliable forecasts of the potential loss of life and societal disruptions during a second wave of soH1N1 in Canada are difficult at best. While initial reports in the Northern Hemisphere have been quite mild and relatively few people are believed to have been infected, there is an anticipation of a stronger fall wave in the Northern Hemisphere during the fall and winter months. Early estimates for the attack rate and case fatality rate (CFR) of the novel H1N1 virus range from 0.0004% to 0.06% (1) though the estimates vary considerable depending upon the method and assumptions used. The early estimates of 0.4% (2) are believed to significantly overestimate the CFR due to poor quality of the early Mexican data. If a significant portion of the population becomes ill, there could be a large number of fatalities both directly due to the virus, and indirectly if the health care system becomes so overburdened with influenza patients that normal levels of care cannot be maintained.

In order to support those responsible for health and economic pandemic response plans for the upcoming flu season, there is a role for dynamic modeling of a pandemic as it unfolds with frequent updates and new parameters estimates as more data becomes available. As a response to seeking out timely and robust pandemic estimates that reflect recent data, a North

American infectious disease model had been constructed to simulate the spread of soH1N1. The model is an advanced SEIR implementation which partitions the population by age, gender and region. The location of first infection is taken to be in Mexico as the soH1N1 virus is believed to have originated in there (3). The model is calibrated against the reported soH1N1 confirmed cases, mortalities, and hospitalizations. Since June 6th 2009, data from the Canadian FluWatch website (4) was used to update the model on a fortnightly basis.

Method

The basic model is an SEIR infectious disease model (5; 6 for example). It is extended to allow age groups, gender, and multiple interacting regions. Technical details are provided in an online appendix A. As the focus of the study is to model the Canadian pandemic potential, the spatial domain is decomposed into 12 Canadian regions (each province, the Yukon, and a single region for the combination of the Northwest Territories and Nunavut), a single region for the United States, and a single region for Mexico. Time-dependent contact rates are also implemented to model school closures during the summer months since, during the summer months, a reduction in contacts among school-aged children may give rise to a temporary suppression in the infection rate. A simple step reduction in contacts among children with their peers is assumed to occur on June 20th, approximately when the school year ends in North America, and return to normal on September 4th when school returns. Key model outcomes include the attack rate defined as the total number of infections divided by the initial population, and the case fatality rate defined as the total number of deaths divided by the total number of infections. Both the attack rate and case fatality rate can also be defined for sub-populations such as specific age groups or gender. In particular, age-dependent attack rates and CFRs can be used to indicate which age groups are at higher risk of infection and death. Due to the lack of publicly

available age-dependent hospitalization data, the case hospitalization rate (CHR) is assumed to be age-independent. In addition, since one is not hospitalized immediately upon infection, the number of people hospitalized is assumed to lag the infection curve. The parameters fit consist of the infectiousness of the novel H1N1 virus, the average infectious period, the case fatality rate, the hospitalization rate, the time of first infection, and the offsets between initial infection and hospitalization or mortality.

The contacts rates used are from Mossong (7) with age-dependent probability of infection from Haber et al. (8). We assume a latency period of 1 day since the quality of fit is quite insensitive to this parameter. At the moment, the CFR is assumed to be age-independent. As more data becomes available, it may be possible to generate age-dependent case fatality rates.

The metric to measure the quality of fit is defined to be

$$\Delta^2 = \sum_{a,r,t} \left(Q_{model}^{a,r,t} - Q^{a,r,t} \right)^2 + w_D \sum_{a,r,t} \left(M_{model}^{a,r,t} - M^{a,r,t} \right)^2 + w_H \sum_{a,r,t} \left(H_{model}^{a,r,t} - H^{a,r,t} \right)^2$$

where $Q^{a,r,t}$ is the cumulative number of confirmed cases, $M^{a,r,t}$ is the cumulative number of H1N1 mortalities, and $H^{a,r,t}$ is the number of hospitalizations all at time t , age a , and region r .

w_D and w_H are the relative weights to give deaths and hospitalizations with respect to confirmed cases. As the data is somewhat sparse and not all regions report at the same time, any times which are missing data are simply omitted from the sums. In addition, since the WHO has decided to no longer track confirmed cases, only confirmed case data from before July 1, 2009 are considered in the first term. The Nelder-Mead simplex algorithm (9) was used to perform a multi-dimensional minimization to find the pandemic parameters that best match the observed

data. The later part of the pandemic fits is then governed by the behaviour of mortalities and hospitalizations rather than the confirmed cases.

An estimate of the base reproductive number, R_0 , can be determined from the model by inserting one infectious individual into each population group, disabling the transition from infected to infectious and calculating the weighted average number of secondary infections (10).

Results

A critical parameter to estimate is the initial detection rate defined as the fraction of people who are confirmed to have the virus out of all people who actually have the virus. A lower bound of the detection rate of the novel H1N1 virus would correspond to the typical confirmation rate for a severe seasonal flu. Using data from the FluWatch website (4), this can be estimated to be approximately 1/300. For example, if one assumes a 10% seasonal flu attack rate in 2004, there would be about 3 million flu cases, but only around 10,000 positive confirmations. Due to increased public awareness and testing, it was assumed that the confirmation rates in the early stages of the pandemic outbreak were significantly better than rates for the seasonal flu so a ratio of 1 in 30 was used. This agrees with early estimates from Fraser et al. (2) and Wilson and Baker (1) for example.

In order to probe the reliability of the method, fits are run using data approximately every two weeks. For trials with data ending prior to June 20th, it is assumed that contacts among children are reduced by 75%. For later dates, the factor reduction in contacts is a fit parameter. Table 1 summarizes the resulting pandemic parameters, while Table 2 shows the calculated R^2 values for each case. For the early fits, the hospitalization data was quite sparse so accurate fits were difficult. However, more recent fits match the observations well. Figure 1 compares the

observed confirmed cases to the modeled values. Observed values beyond July 1 are ignored due to the changes in testing policy and are therefore not expected to match the modeled values.

Figures 2 and 3 compare the observed number of cumulative mortalities and hospitalizations to the modeled values respectively. Due to relatively little hospitalization data during the early stages of the pandemic, the initial hospitalization curves tend to vary considerably as new data is added, but more recent fits have stabilized. Figure 4 shows the expected number of infectious people in Canada over the course of each modeled pandemic. While the location of the peak in each curve does vary, the definite trend is for the pandemic to peak in late fall or early winter of 2009.

The attack rate is strongly dependent upon the age group in question. Figure 5 shows the attack rates in 5-year age groups for the fit using data until September 5th. The age-dependence is driven by the contact rates as the observed data had no age discrimination. Younger people, particularly those between 10 and 20, have much higher contact rates and therefore more opportunities to come in contact with an infected individual. This variation in attack rate must be taken into account when estimating the hospitalization rates and associated risks. For example, there could be many more 15 to 20 year olds hospitalized compared to 65 to 70 year olds, but the risk of hospitalization given infection could still be much larger for the older age group. Age-dependent data would be required to accurately determine the age-dependent risks of hospitalization and mortality.

Discussion

The pandemic fit algorithm developed currently fits the observed data extremely well. The downturn in infections and mortalities seen over the late summer was successfully accounted for by a reduction in contacts among school-aged children during the summer months. This implies that the spread of the soH1N1 virus will accelerate once school resumes, and initial reports from regions where school has begun support this supposition (11).

There are limitations and caveats to the estimations presented. The objective of the modeled fits is not to provide a precise prediction of the future, but rather to provide knowledge transfer and knowledge translation services to pandemic responders in the midst of a pandemic. As to the knowledge transfer role, the mortalities and hospitalizations in the short term can be used to estimate upcoming potential demand while the longer term forecast indicates the possible trends. As more data becomes available and other factors such as changes in social behaviour and vaccination start to play a role, the model's long term behaviour could change significantly. The significance of change will not be known until new data is introduced with characteristics that had not been anticipated by the model. Such changes, if they were to occur, would be considered useful information for pandemic responders as the model flags the introduction of new information (as opposed to updated information that is fully anticipated by the model). In this way, dynamic modeling of this kind serves a knowledge transfer role for pandemic responders, as the introduction of new data in the midst of a pandemic is transferred to future forecasts of potential pandemic impact and resource capacity significance. As to the knowledge translation role of the approach, the model provides a framework in which a dynamic and iterative process that synthesises several inputs (eg. disease, population, logistic) can occur.

As more data continues to become available, the fit to the model can continue to be improved and updated forecasts generated. The weaknesses of any model have much to do with

the assumptions that go into it. If the model can no longer fit the data reasonably well, it is a signal that additional processes must be integrated into the model. For example, when vaccinations become available, the immunization timetable of the high priority groups and the vaccine's efficacy will have to be added to the model. If the model successfully can track the pandemic up to the point of vaccination, the model can also be used as a tool to analyze how the predicted vaccination efficacy compares to the observed behaviour and vaccine strategies could be adjusted accordingly if required.

References

1. **Wilson, N and Baker, M G.** The Emerging Influenza Pandemic: Estimating the Case Fatality Ratio. *Euro Surveill.* July 2, 2009.
2. **Fraser, Christophe et al.** Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Scienceexpress.* May 11, 2009.
3. **Smith, G. J., et al.** Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature.* June 25, 2009, Vol. 459, 7250, pp. 1122-5.
4. **Public Health Agency of Canada.** FluWatch. [Online] 2009. [Cited: September 10, 2009.] <http://www.phac-aspc.gc.ca/fluwatch/index-eng.php>.
5. **Hall, I. M., et al.** Real-time epidemic forecasting for pandemic influenza. *Epidemiol Infect.* April 2007, Vol. 135, 3, pp. 372-385.
6. **Flahault, Antoine, Deguen, Séverine and Valleron, Alain-Jacques.** A mathematical model for the European spread of influenza. *European Journal of Epidemiology.* August 1994, Vol. 10, 4.
7. **Mossong, Joel, et al.** Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine.* March 25, 2008.
8. **Haber, Michael J, et al.** Effectiveness of Interventions to Reduce Contact Rates during a Simulated Influenza Pandemic. *Emerging Infectious Diseases.* April 4, 2007. <http://www.cdc.gov/EID/content/13/4/581.htm>.

9. **Nelder, J. A. and Mead, R.** A simplex method for function minimization. *Computer Journal*. 1965, Vol. 7, pp. 308-313.

10. **Heffernan, J. M., Smith, R. J. and Wahl, L. M.** Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*. 2005, 2, pp. 281-293.

11. **The Globe and Mail.** Hundreds stay home as flu hits schools before vaccine is ready. *The Globe And Mail*. [Online] September 24, 2009. [Cited: October 7, 2009.]
<http://www.theglobeandmail.com/news/national/three-bc-schools-report-students-sick-with-h1n1-flu-virus/article1299017/>.

Biographical Sketch for Paul Smetanin

Paul Smetanin is the CEO and President of RiskAnalytica and is a doctoral candidate at the Graduate College of Management, Southern Cross University, East Lismore, NSW, Australia, postal code 2480. Paul has a Masters of Quantitative Methods and a Bachelor of Economics.

Table 1: Summary of the fit results using data up to the given dates.

Date of Fit	Attack Rate (%)	Infectious Period (days)	CFR (%)	Date of Second Wave Peak	R_0	Total Mortalities	CHR (%)
2009-06-06	25.8%	2.49	0.030%	2009-11-29	1.97	2613	0.20%
2009-06-20	21.4%	2.33	0.034%	2009-12-07	1.91	2442	0.11%
2009-07-04	27.5%	2.42	0.030%	2009-12-11	1.94	2733	0.20%
2009-07-13	26.9%	2.25	0.046%	2010-01-01	1.95	4140	1.18%
2009-07-31	26.9%	2.53	0.017%	2009-12-25	1.96	1504	0.41%
2009-08-17	23.2%	2.37	0.018%	2009-12-29	1.94	1367	0.36%
2009-09-05	29.2%	2.33	0.019%	2009-12-25	1.94	1855	0.42%

Table 2: Summary of quality of fit (R^2 values) using data up to the given dates.

Generally, as more data became available, the quality of fit improved.

Date of Fit	Confirmed Cases	Mortalities	Hospitalizations
2009-06-06	0.99005	0.77480	NA
2009-06-20	0.99562	0.82786	-0.27996
2009-07-04	0.99674	0.94497	-0.39273
2009-07-13	0.99731	0.92738	0.93448
2009-07-31	0.99546	0.98569	0.98993
2009-08-17	0.99782	0.99061	0.99241
2009-09-05	0.99767	0.99172	0.98810

Figure 1: Comparison between observed confirmed cases and modelled confirmed cases using data up to the dates specified. The models diverge from the observed after June due the changes in confirmed case reporting.

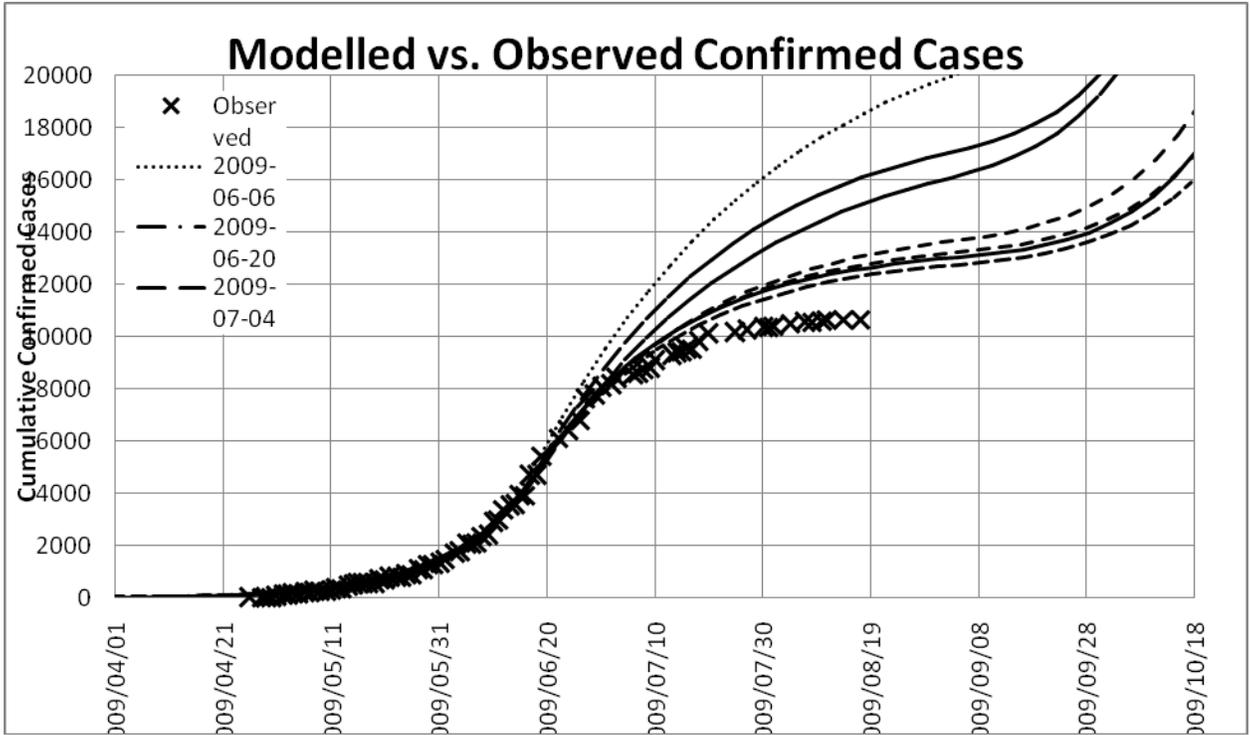


Figure 2: Comparison between observed mortalities and modelled mortalities using data up to the dates specified for each curve.

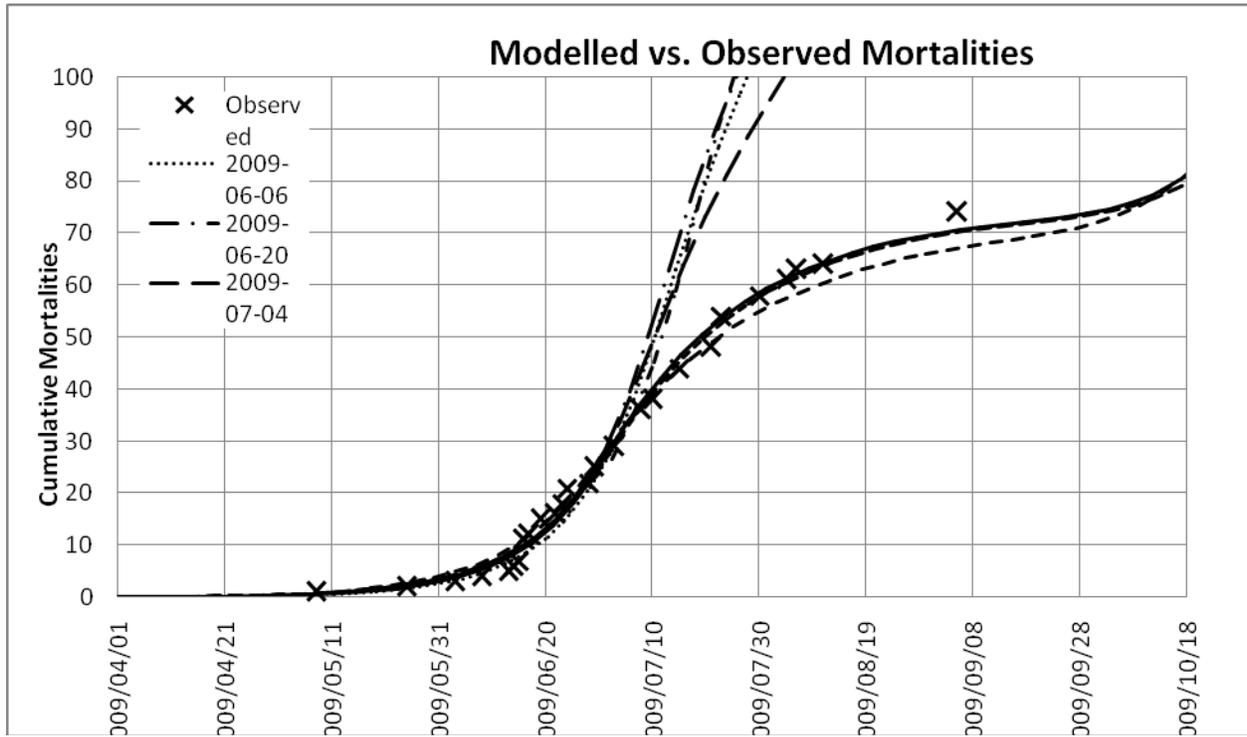


Figure 3: Comparison between observed hospitalizations and modelled hospitalizations using data up to the dates specified. Hospitalization estimates based on early data tend to fit poorly due to the small number of reported cases.

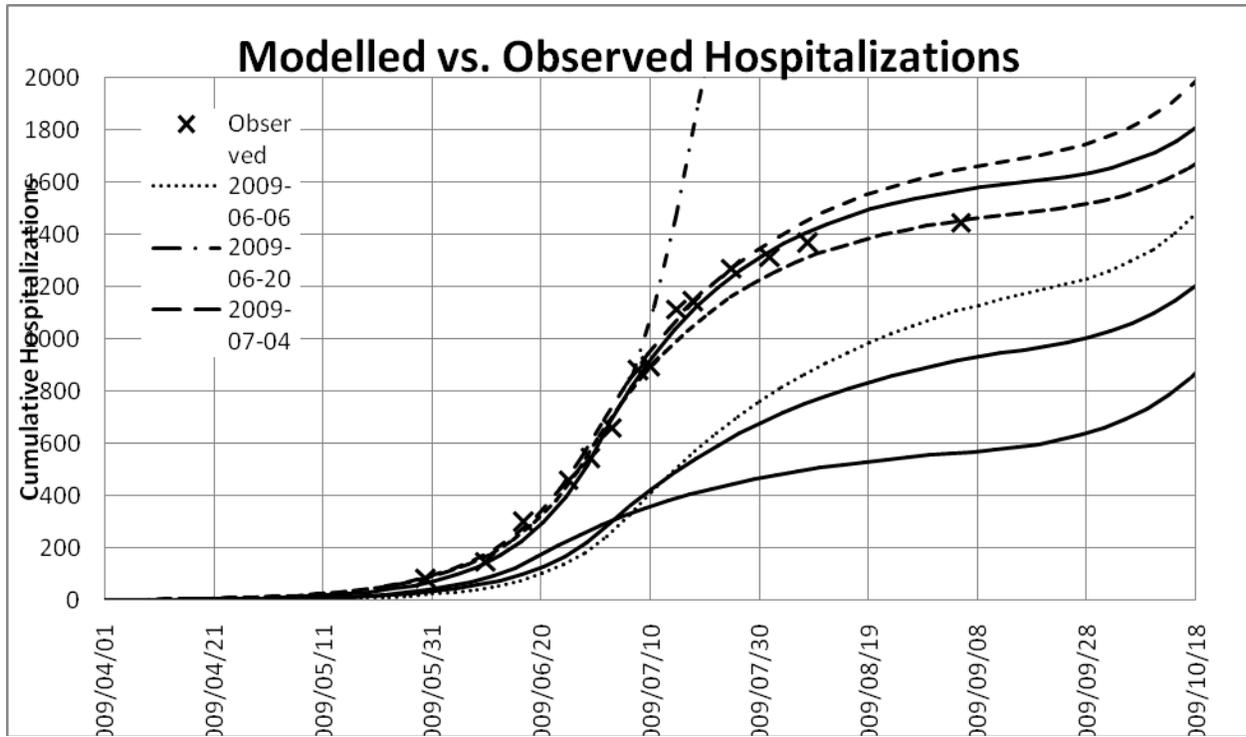


Figure 4: The infectious population that results from the fits. The early wave dies out due to changes in contact patterns over the summer but resumes when school returns. The peak number of infections tend to be around late fall or early winter.

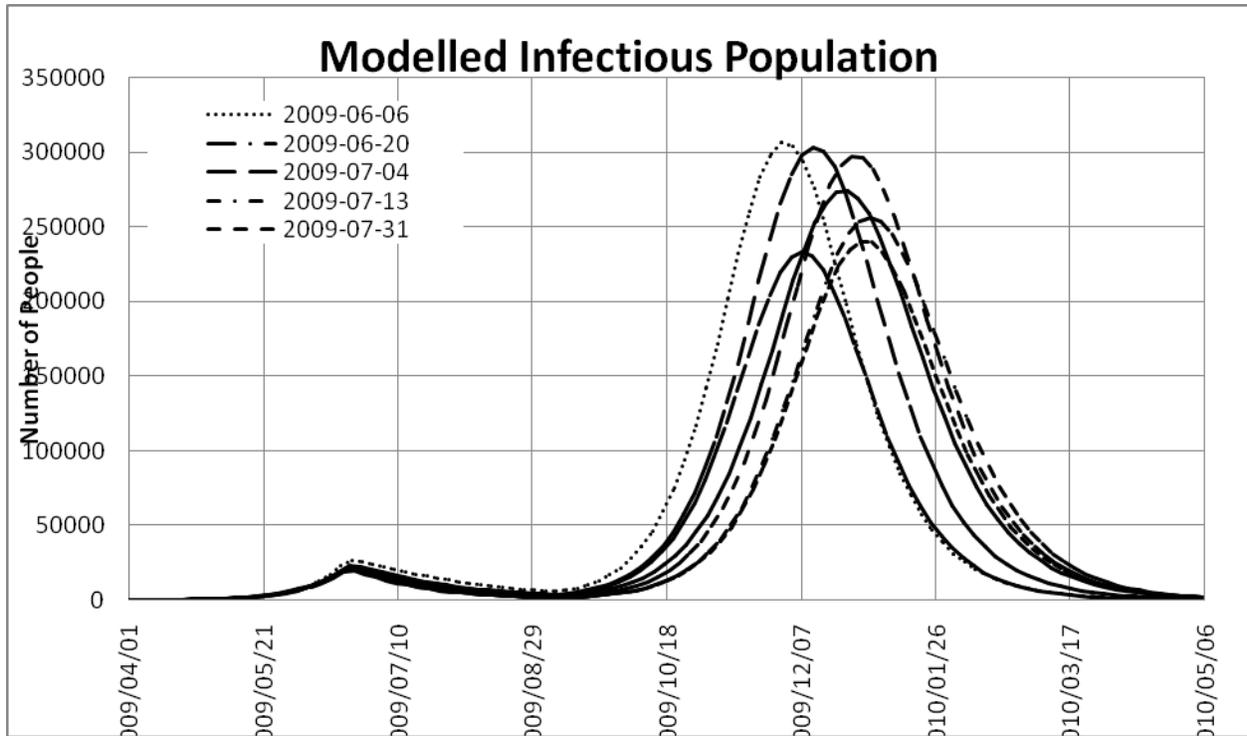
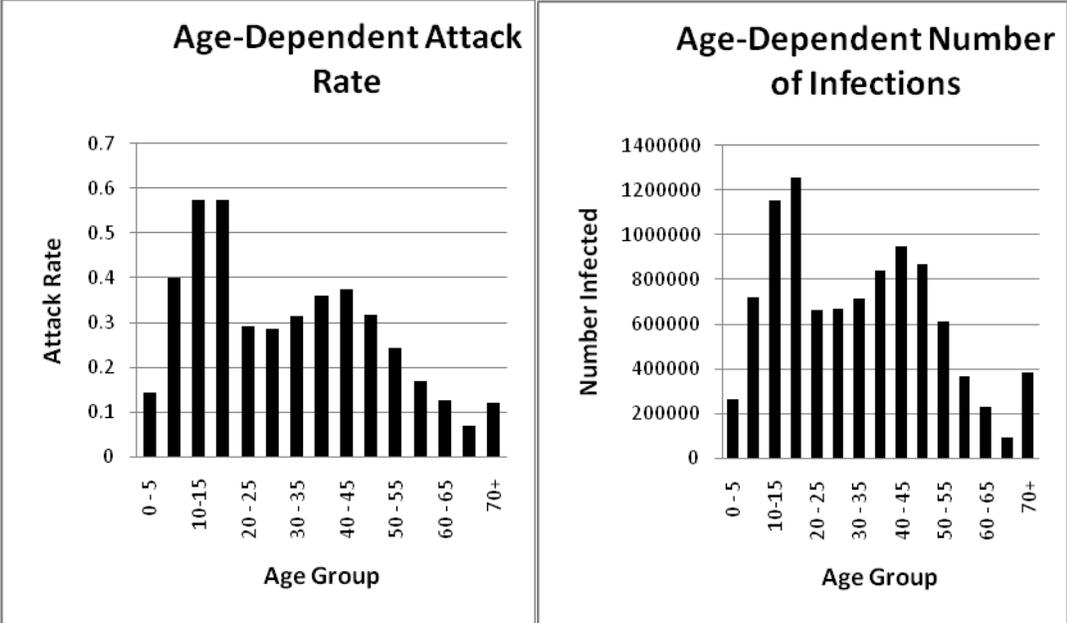


Figure 5: Age-dependent attack rates (left) and total infections (right) for the September 5th fit.



Technical Appendix

The model equations are

$$\begin{aligned}
 \frac{dS_r^{ag}}{dt} &= -S_r^{ag} \underbrace{\sum_b C_r^{ab} \rho^{ab} F_r^{abg}}_{\text{Infection}} \\
 \frac{di_r^{ag}}{dt} &= S_r^{ag} \underbrace{\sum_b C_r^{ab} \rho^{ab} F_r^{abg}}_{\text{Infection}} - \underbrace{\lambda_{i \rightarrow I} i_r^{ag}}_{\text{Latent period}} \\
 \frac{dI_r^{ag}}{dt} &= \underbrace{\lambda_{i \rightarrow I} i_r^{ag}}_{\text{Latent period}} - \underbrace{\lambda_{I \rightarrow R} I_r^{ag}}_{\text{Recovery}} - \underbrace{\lambda_{I \rightarrow D} I_r^{ag}}_{\text{Death}} \\
 \frac{dR_r^{ag}}{dt} &= \underbrace{\lambda_{I \rightarrow R} I_r^{ag}}_{\text{Recovery}}
 \end{aligned}$$

where S_r^{ag} , i_r^{ag} , I_r^{ag} and R_r^{ag} are the susceptible, infected, infectious, and recovered populations respectively in age group a , of gender g , and in region r . The rates of recovery, death and transition from infected to infectious are $\lambda_{I \rightarrow R}$, $\lambda_{I \rightarrow D}$, and $\lambda_{i \rightarrow I}$ respectively. ρ^{ab} is the probability of infection given that a person in age group a has come into contact with an infectious individual of age group b . C_r^{ab} is the rate that people age a contact people of age b . The quantity F_r^{abg} , described below, is the effective fraction of contacts a person in region r of age a makes with infectious persons of age b which includes the impact of travel between regions.

In order to couple the various regions in the model, travel data from Canadian and international travel surveys (1; 2; 3) was used to calculate the fraction of the population from each region which would currently be in any other region. The regional coupling coefficient, ϵ_{rk}^{ag} , indicates the fraction of the population of age a and gender g who normally reside in region r that are currently in region k . The effective infectious population and total population in region r are then $I_{\text{eff},r}^a = \sum_{d,g} \epsilon_{dr}^{ag} I_d^{ag}$ and $P_{\text{eff},r}^a = \sum_{d,g} \epsilon_{dr}^{ag} P_d^{ag}$ respectively. It is also possible that a susceptible person has left region r and undergone an encounter with an infectious individual in region k . Therefore, the final effective fraction of contacts that are infectious is the weighted average of the possible locations where the susceptible person may be located and is defined as

$$F_r^{ab,g} = \sum_d \epsilon_{rd}^{ag} \frac{I_{\text{eff},d}^b}{P_{\text{eff},d}^b}$$

It is useful to introduce the cumulative number of infections by age, gender and region, Q_r^{ag} , and the cumulative number of deaths D_r^{ag} . Their differential equations are

$$\frac{dQ_r^{ag}}{dt} = S_r^{ag} \underbrace{\sum_b C_r^{ab} \rho^{ab} F_r^{abg}}_{\text{Infection}}$$

$$\frac{dD_r^{ag}}{dt} = \underbrace{\lambda_{I \rightarrow D} I_r^{ag}}_{\text{Death}}$$

The attack rate is then defined as

$$A = \frac{Q(t_\infty)}{P(t_0)}$$

where t_0 is the time of initial infection, and t_∞ is some time after the pandemic has passed. The case fatality rate is the fraction of people who were infected that died and is defined

$$CFR = \frac{D(t_\infty)}{Q(t_\infty)}$$

Both the attack rate and case fatality rate can also be defined for sub-populations such as specific age groups or gender. In particular, age-dependent attack rates and CFRs can be used to indicate which age groups are at higher risk of infection and death.

Due to the lack of publicly available age-dependent hospitalization data, the case hospitalization rate (CHR), α_h , is assumed to be age-independent. In addition, since one is not hospitalized immediately upon infection, the number of people hospitalized, H^{ag} , is assumed to lag the infection curve.

$$\frac{dH^{ag}}{dt} = \alpha_h \frac{dQ^{ag}}{dt} (t - \tau_H)$$

Finally, as the model does not take into account the effect of hospitalization and critical care delaying mortalities, an offset between “model death” and observed mortalities, M^{ag} , is also introduced.

$$M^{ag}(t) = D^{ag}(t - \tau_D)$$

The parameters fit consist of the infectiousness of the novel H1N1 virus, the average infectious period, the case fatality rate, the hospitalization rate, the time of first infection, and the offsets between initial infection and hospitalization or mortality.

An estimate of the base reproductive number, R_0 , can be determined from the model by inserting one infectious individual into each population group, disabling the transition from infected to infectious ($\lambda_{i \rightarrow I} = 0$) and calculating the weighted average number of secondary infections (4).

$$R_0 = \frac{\sum_{a,g,r} P_r^{ag}(t_0) Q_r^{ag}(t_\infty)}{\sum_{a,g,r} P_r^{ag}(t_0)}$$

References

1. **Statistics Canada.** *International Travel Survey.* 2007. Catalogue no. 66-201-X..
2. **US Bureau of Transportation Statistics.** *US-International Travel and Transportation Trends: 2006 Update.* 2006.
3. **Statistics Canada.** *Canadian Travel Survey.* 2004.
4. **Heffernan, J. M., Smith, R. J. and Wahl, L. M.** Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface.* 2005, 2, pp. 281-293.
5. **Mossong, Joel, et al.** Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine.* March 25, 2008.
6. **Haber, Michael J, et al.** Effectiveness of Interventions to Reduce Contact Rates during a Simulated Influenza Pandemic. *Emerging Infectious Diseases.* April 4, 2007.
<http://www.cdc.gov/EID/content/13/4/581.htm>.
7. **Hall, I. M., et al.** Real-time epidemic forecasting for pandemic influenza. *Epidemiol Infect.* April 2007, Vol. 135, 3, pp. 372-385.