Patterns of spread of influenza A in Canada

Daihai He1, Jonathan Dushoff2,3, Raluca Eftimie4 and David J. D. Earn3,5

1Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (SAR), People's Republic of China
2Department of Biology, and 3M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada L8S 4L8
4Division of Mathematics, University of Dundee, Dundee DD1 4HN, UK
5Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1

Understanding spatial patterns of influenza transmission is important for designing control measures. We investigate spatial patterns of laboratory-confirmed influenza A across Canada from October 1999 to August 2012. A statistical analysis (generalized linear model) of the seasonal epidemics in this time period establishes a clear spatio-temporal pattern, with influenza emerging earlier in western provinces. Early emergence is also correlated with low temperature and low absolute humidity in the autumn. For the richer data from the 2009 pandemic, a mechanistic mathematical analysis, based on a transmission model, shows that both school terms and weather had important effects on pandemic influenza transmission.

1. Introduction

Seasonality of transmission rates is known to have a significant influence on the temporal patterns of epidemics of infectious diseases [1]. Transmission rates appear to be influenced both by contact patterns (e.g. aggregation of children in schools when schools are in session [2–5]) and by weather patterns (e.g. changes in humidity and/or temperature [5–10]). Understanding how these factors—and others, such as travel patterns—affect the full spatio-temporal dynamics of infectious disease spread is important for epidemic prediction and control.

The spatial spread of influenza across countries and continents has become a very active area of research [11–15], and recent work has begun to connect transmission mechanisms with observed spatial patterns. In particular, influenza mortality patterns across the continental United States are correlated with differences in absolute humidity [7]. Here, we investigate the relationships among weather variables (temperature and humidity), school calendars and influenza incidence across the 10 Canadian provinces over the last 14 years. We use statistical models to investigate correlations over this period, and in the case of the 2009 pandemic, we also fit mechanistic mathematical models that allow us to draw stronger conclusions about the effects of seasonal factors on influenza transmission and incidence.

2. Material and methods

(a) Data sources

Weekly influenza A laboratory-confirmed cases between 23 October 1999 and 9 March 2013 were taken from the Public Health Agency of Canada FluWatch surveillance programme reports [16]. Daily climate data were obtained from Canada’s National Climate Archive [17]. We used climate data for the most populous city in each province to represent the province. School opening and closing dates were obtained from the Canadian Education Association (http://www.cea-ace.ca/); if these dates were not uniform throughout a province, we averaged the published dates.

Population sizes and estimated pandemic H1N1 (pH1N1) vaccination coverage in each of the 10 Canadian provinces in 2009 and 2010 were obtained from Statistics
Canada [18,19]. In Toronto (Ontario, ON) and Montreal (Quebec, QC), the majority of vaccine doses were delivered during November and December [20,21]. Since vaccination dates were not available for other provinces, we assumed that they also vaccinated during the same time span.

(b) Statistical analysis: generalized linear model
We looked for patterns in influenza epidemic timing for the 11 seasonal influenza (nine pre-pandemic and two post-pandemic) seasons in our dataset. Each flu season starts from the 35th week of each year and testing occurs throughout the year in Canada [16]. To avoid spurious findings, we restricted ourselves to a single set of predictors specified a priori in the main analysis:

- spatial ordering of provinces from east to west (longitude rank),
- population rank of the provinces (population rank),
- mean observed October humidity, and
- mean observed October temperature.

To facilitate an unambiguous ranking of longitude (see geographical locations of the 10 Canadian provinces in the electronic supplementary material), we grouped Canada’s four Atlantic provinces into one ‘mega-province’. Thus, our analysis treated seven regions of Canada (six provinces and one mega-province).

To test the significance of our hypotheses, we used the R statistical programming language (http://www.R-project.org/) to model epidemic timing as a function of the four predictors. We also considered influenza season (treated categorically) to account for the fact that overall timing was different in different years. Our measure of epidemic timing was the date on which the number of cumulative cases reported reached a given proportion of the seasonal total for a given season in each province, which we call the quantile time. To test whether a set of predictors is significant as a group, we used an analysis of variance test to compare the original model to a model that excludes the focal predictors.

(c) Mathematical model and data fitting
To connect seasonal factors with influenza transmission during the autumn wave of the 2009 pandemic (4 July 2009 to 18 December 2009), we used a standard mechanistic mathematical model, the classical SIR model [1]

\[
\dot{S} = -\beta(t)SI - v(t),
\]

\[
\dot{I} = \beta(t)SI - \gamma I - (1 - \rho)\gamma I.
\]

Here, S and I represent the numbers of susceptible and infectious individuals, γ is the recovery rate (the rate at which individuals move from the infectious class to the recovered class), and we assumed that the mean infectious period (1/γ) is 3 days [5,22–25]. ρ is the reporting ratio (the proportion of cases that are reported). v(t) is the vaccination rate. We define

\[
v(t) = \begin{cases} \frac{\eta \delta S(0)}{D}, & t \in (10 \text{ October 2009, 10 January 2010}) \quad (2.2) \\ 0, & \text{otherwise.} \end{cases}
\]

where η denotes the vaccine effectiveness, η = 66% [26]; δ denotes the vaccination coverage and the estimation is available for each province [19]; S(0) denotes the initial susceptibles; D denotes the duration of the vaccination campaign, which was approximately two months. The vaccination campaign was started in the last week of October 2009 [20]. We assume that the vaccination became effective between (10 October 2009 and 10 January 2010), to allow for the delay in vaccine-induced protection [26].

Testing was restricted for a substantial portion of the 2009 pandemic, so the reporting ratio ρ was not constant throughout the epidemic. However, restrictions were initiated before our analysis period started in July [27], allowing us to use a constant ρ—with the exception of Alberta, where testing was not restricted until November [5]. Hence we used a step function for ρ in Alberta.

In equations (2.1a,b), the transmission rate β is time-varying (dependent on weather and the school calendar). We took it to have the following specific form:

\[
\beta(t) = \beta_0 \cdot e^{-a_HH(t)/T(t)} \cdot e^{-a_TT(t)} \times \begin{cases} 1, & \text{summer vacation.} \\ 1 + e, & \text{school term.} \end{cases}
\]

(2.3)

Here, H(t) is the absolute humidity and T(t) is the temperature at time t. The parameter β0 is the baseline transmission rate during the summer vacation. The parameter e, which we call the school term intensity, controls the increment in transmission after schools reopen in September. The parameters α_H and α_T describe the strength of response to humidity and temperature, respectively. This definition of β(t) gives us 2^3 = 8 scenarios to consider, since we can separately turn off the effects of school terms, temperature or humidity. Below we focus on the following three ‘sub-models’: (i) only school terms and temperature (α_H = 0), (ii) only school terms and humidity (α_T = 0) and (iii) only school terms (α_H = α_T = 0).

Because temperature and humidity are highly correlated, and the functional responses are imperfectly known, we do not expect to be able to disentangle the separate contributions of these two factors. We therefore do not analyse any models that incorporate both humidity and temperature.

In our simulations, we assumed that the initial proportion of susceptible individuals S(0)/N was 65% (on 4 July 2009; in the electronic supplementary material, we investigate two other assumptions for this quantity), and the initial proportion of infected individuals was below 1% (cf. [28–31]). We define the basic reproduction number \( R_0 = (\beta(t))/\gamma \) [32], where the expectation is taken over the time window that we model. We also define the effective reproduction number \( R_e = R_0 \frac{S(0)}{N} \).

3. Results
(a) Patterns of annual influenza epidemic spread across Canada
Figure 1a shows the weekly laboratory-confirmed influenza A cases for each Canadian province from October 1999 to March 2013 (the last few months are not included in our analysis, since we do not have complete data for this flu year). The top axis indicates which subtype was most common in each season. Epidemics tended to begin earlier in the H3N2-dominated seasons. In addition, three provinces (British Columbia (BC), Alberta (AB) and ON) tended to have earlier influenza epidemics than the rest of Canada.

Figure 1b–d shows the times at which 25, 50 and 75% of weekly laboratory-confirmed cases occurred in the non-pandemic years (i.e. excluding the 2009–2010 and 2010–2011 flu years). Figure 1b suggests that on average, influenza tends to start earliest in AB, followed by BC and ON, eventually reaching the Maritimes (MA) last. There is much less evidence of spatial spread when we look later in the seasons (figure 1c,d). Figure 1e shows the temperature pattern, which may explain why flu starts in BC later than in AB; humidity (not shown here) follows a similar pattern. We use the Tukey-
Spatio-temporal pattern of influenza A transmission in Canada. (a) Weekly laboratory-confirmed influenza A cases in the 10 Canadian provinces (per 100,000 inhabitants) from October 1999 to March 2013. Labels at the top show the dominant subtype(s) in Canada in each influenza season [16]. Vertical dashed lines are shown on 1 January and 1 July of each year. (b–d) Deviations of the provincial quantile times (at 25, 50 and 75%) from corresponding Canadian quantile times for each year. MA here refers to the Maritime provinces of NB, NS, PE and NL. (e) The number of days (between September and December, for each year between 1999 and 2012) when the mean temperature was below zero. Provinces are ordered in longitude order from west to east (see the electronic supplementary material, figure S4). (Online version in colour.)

Figure 1. Spatio-temporal pattern of influenza A transmission in Canada. (a) Weekly laboratory-confirmed influenza A cases in the 10 Canadian provinces (per 100,000 inhabitants) from October 1999 to March 2013. Labels at the top show the dominant subtype(s) in Canada in each influenza season [16]. Vertical dashed lines are shown on 1 January and 1 July of each year. (b–d) Deviations of the provincial quantile times (at 25, 50 and 75%) from corresponding Canadian quantile times for each year. MA here refers to the Maritime provinces of NB, NS, PE and NL. (e) The number of days (between September and December, for each year between 1999 and 2012) when the mean temperature was below zero. Provinces are ordered in longitude order from west to east (see the electronic supplementary material, figure S4). (Online version in colour.)

Using a generalized linear model controlling for the effect of variation among influenza seasons (see Material and methods), we analysed the roles of four predictors: longitude rank, population rank, humidity and temperature. We are mostly interested in the initial spreading, thus we used 25% as the threshold proportion (we show other cases in the electronic supplementary material). We found that all four predictors as a group are significant (\(p \approx 0.001\)). In particular, longitude rank, population rank, humidity and temperature are significant individually (\(p < 0.001, 0.021 \) and 0.016, respectively). Although population is not a significant predictor (\(p = 0.413\)), a direct comparison between longitude rank and population rank does not show that the former is significantly better than the latter.

(b) Pattern of 2009 pandemic spread across Canada

Recent work has indicated that school closures lead to a significant reduction in transmission of influenza (e.g. seasonal influenza in France [33] and pandemic influenza in the province of AB, Canada [5]). Figure 2 suggests that summer closings of schools in 2009 influenced the transmission dynamics of pandemic influenza throughout Canada and led to substantially fewer cases during the summer months than would have occurred if schools had remained open all summer. The figure reveals a variety of suggestive patterns: for example, provinces that closed schools earlier (e.g. QC) seem in general to have experienced fewer cases, whereas provinces that opened schools earlier (e.g. AB and Saskatchewan (SK)), seem to have experienced an earlier second wave. However, the pattern of autumn spread also seems broadly consistent with the spatio-temporal patterns of temperature and absolute humidity; in particular, a wave of low temperature (and absolute humidity) in October 2009 is correlated with the peak of the second wave of the pandemic. Also worth noting is the relatively low number of reported cases in the more populous provinces of ON and QC (probably due to testing restrictions) and many periods with very few reports from Manitoba (MB), Nova Scotia (NS), Prince Edward Island (PE) and Newfoundland (NL) between April and October 2009 (possibly indicating local fade-out and re-introduction).

To explore these patterns mechanistically, we fit an SIR model to the pandemic data (see Material and methods). Figure 3 shows the log-likelihood profiles of the three sub-models as a function of school term intensity (\(e\), equation (2.3)) in each of the seven regions of Canada (six provinces and the maritime ‘mega-province’). Higher likelihood indicates better fit. It is evident that sub-model (i), with an additional effect of temperature variation, is the best among the three sub-models in all regions. Sub-model (ii), with humidity replacing temperature, is as good as sub-model (i) in ON and MB. Sub-model (iii), with neither weather variables, is clearly the worst in all regions. Note that sub-models (i) and (ii) involve the same number of parameters, while sub-model (iii) involves one fewer parameter. The difference between sub-model (iii) and sub-models (i) and (ii) is much more than can be explained by reducing the number of parameters by one (\(\approx 4\) units of log likelihood). Electronic supplementary material, table S2 shows the second-order Akaike information criterion (AIC\(_c\)) [5,8,35] for the three sub-models for each of the seven regions (AIC\(_c\) formalizes a tradeoff between model complexity and goodness of fit).
**Figure 2.** Spatial pattern of influenza A (mainly pandemic H1N1) transmission in Canada (a), as well as daily mean absolute humidity (b) and temperature (c) across all Canadian provinces, between September 2008 and July 2010. (a) Also shows school closing dates (the thin vertical lines in late June) and opening dates (the thick vertical lines in early September). In some provinces, there were several school closing/opening dates, corresponding to different regions and/or for different grades. (Online version in colour.)

**Figure 3.** Likelihood profiles of three sub-models as a function of school term factor in each of the seven Canadian regions. The likelihood profile is obtained by maximizing the likelihood of each of the three sub-models while fixing the school term factor at values spanning from 0.01 to 2.3, via iterated filtering [34]. Three symbols (circle (temperature sub-model), square (humidity sub-model) and triangle (no-weather sub-model)) correspond to sub-models (i) – (iii) as defined in Material and methods. Sub-model (i), with an effect of temperature variation, is the best. Sub-model (iii), with neither weather factors, is the worst in all provinces. Comparison of their second-order Akaike information criterion is given in the electronic supplementary material. (Online version in colour.)
Figure 3 also reveals a relationship between the effects of weather and school term in the data. When the school term intensity is large, the difference between sub-models with and without weather decreases; whereas when the school term intensity is small, sub-models that include weather fit the data much better than those that include neither weather variable. These observations suggest that the effects of school term and weather variation are similar. We also note that the likelihood profiles for the prairie provinces (AB, SK and MB) have similar shapes, perhaps due to the proximity in geography and similarity in climate among these provinces. The 95% CIs for the school term intensity can be calculated from figure 3 via the procedure described in [36].

Table 1 shows our maximum-likelihood estimates of parameters: the reproductive number ($R_e$), the attack rate (AR), the total infected proportion), the baseline transmission rate ($\rho$), the temperature intensity and the school term intensity ($\varepsilon$), in each of the seven regions, based on sub-model (i). The $R_e$ estimates for the autumn pandemic wave are close to published estimates for the spring wave [5,23–25]. Our estimated ARs are larger than published values, e.g. 36% published in [37] (for the whole pandemic including vaccination). However, the published values were based on haemagglutination inhibition (HI) titres greater than or equal to 40; if a lower threshold (greater than or equal to 20) is used, the estimated AR is larger [30,37].

Electronic supplementary material, figures S9 and S10, compare the observed data with simulations with estimated parameter values for the three sub-models in each of the seven regions. It can be seen that simulations without weather variables match the observed data less well than those that include weather variables.

Electronic supplementary material, figure S7, shows total number of laboratory-confirmed influenza cases in each region during the pandemic year (with a box plot of annual cases during pre-pandemic seasonal epidemics for comparison), the reporting ratio and the school term intensity (with 95% confidence interval), estimated from the three sub-models for the pandemic year. The results show that the sub-model without temperature or humidity gives higher estimates of reporting ratio and school term intensity.

### Table 1. The maximum-likelihood estimates of parameters for the autumn wave of the 2009 influenza pandemic in seven Canadian regions, estimated using sub-model (i), the sub-model with an effect of temperature variation. The parameters are the reproductive number, the attack rate (AR), the reporting ratio ($\rho$), the baseline transmission rate, the temperature intensity ($\alpha_T$) and the school term intensity ($\varepsilon$). For $\alpha_T$ and $\varepsilon$, the 95% confidence intervals (CIs) are given in parentheses.

<table>
<thead>
<tr>
<th>province</th>
<th>$R_e$</th>
<th>AR (%)</th>
<th>$\rho$</th>
<th>$\alpha_T$</th>
<th>95% CI of $\alpha_T$</th>
<th>$\varepsilon$</th>
<th>95% CI of $\varepsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>1.27</td>
<td>43.5</td>
<td>0.0035</td>
<td>374.37</td>
<td>(0.033, 0.054)</td>
<td>0.088</td>
<td>(0.010, 0.247)</td>
</tr>
<tr>
<td>AB</td>
<td>1.25</td>
<td>42.6</td>
<td>0.0054</td>
<td>229.48</td>
<td>(0.023, 0.034)</td>
<td>0.247</td>
<td>(0.180, 0.343)</td>
</tr>
<tr>
<td>SK</td>
<td>1.32</td>
<td>45.7</td>
<td>0.0060</td>
<td>229.14</td>
<td>(0.022, 0.039)</td>
<td>0.261</td>
<td>(0.146, 0.484)</td>
</tr>
<tr>
<td>MB</td>
<td>1.32</td>
<td>45.9</td>
<td>0.0033</td>
<td>249.63</td>
<td>(0.013, 0.049)</td>
<td>0.384</td>
<td>(0.083, 0.614)</td>
</tr>
<tr>
<td>ON</td>
<td>1.30</td>
<td>45.0</td>
<td>0.0012</td>
<td>406.49</td>
<td>(0.035, 0.052)</td>
<td>0.075</td>
<td>(0.010, 0.261)</td>
</tr>
<tr>
<td>QC</td>
<td>1.27</td>
<td>45.3</td>
<td>0.0034</td>
<td>368.04</td>
<td>(0.042, 0.053)</td>
<td>0.018</td>
<td>(0.010, 0.189)</td>
</tr>
<tr>
<td>MA</td>
<td>1.31</td>
<td>47.0</td>
<td>0.0044</td>
<td>411.93</td>
<td>(0.058, 0.077)</td>
<td>0.011</td>
<td>(0.010, 0.113)</td>
</tr>
</tbody>
</table>

4. Discussion and conclusion

In this article, we have reported evidence of a west-to-east spatial pattern of spread of seasonal influenza A across Canada between 1999 and 2013, using a simple statistical model. We also studied the autumn wave of the 2009 pandemic in more detail, using a model with explicit transmission, which allowed us to demonstrate that the spread of influenza across Canada is strongly affected by weather even when school terms are accounted for. For seasonal influenza, we demonstrated that longitude rank, humidity and temperature are significant predictors. We chose not to fit a mechanistic model to the seasonal influenza data, since seasonal influenza tends to be sampled more poorly, and the data generally include multiple strains.

Our results on the pattern of seasonal epidemic spread are consistent with an earlier phenomenological report that covered both the USA and Canada [15]. Our results on the importance of weather in pandemic transmission are consistent with similar observations across the United States for seasonal influenza [7].

This study aggregated cases at the provincial level and did not distinguish between age groups. This allowed us to obtain robust fits, but may have led us to miss some details. The aggregated fits also complicate the interpretation of our school term effect. The direct effect of school terms on transmission among schoolchildren will be greater than our estimated effect, averaged over the whole population. Quantifying this difference would require an understanding of the extent to which schoolchildren drive the influenza epidemic, both when schools are open and when they are closed, which in turn would require an age-structured model.

In the electronic supplementary material, we also investigated the timing of intervention and showed its importance on the reduction of infections. While this study is not the first one to investigate vaccination measures in Canada during the 2009 pandemic (see [38]), it is the first that includes the effects of weather. Overall, our study suggests that better understanding of the factors underlying patterns of spatio-temporal spread will be very useful for designing and prioritizing vaccination and other control efforts. As influenza surveillance and modeling techniques continue to improve, it should become possible to further unravel the climate factors that affect influenza transmission using data from seasonal epidemics.

Acknowledgements. We thank two anonymous referees for helpful comments on the manuscript. The authors thank Dr Joseph Wu and Mike Delorme for helpful discussion.

Funding statement. J.D. and DJ.D.E. were supported by grants from CIHR and NSERC. D.H. was supported by a start-up grant from the Department of Applied Mathematics at Hong Kong Polytechnic University (HKP) and an HKP ‘Central Bidding’ grant.