Outpatient physician billing data for age and setting specific syndromic surveillance of influenza-like illnesses

Emily H. Chan\textsuperscript{a}, Robyn Tamblyn\textsuperscript{a}, Katia M.L. Charland\textsuperscript{a,b}, David L. Buckeridge\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}McGill University, Montreal, Quebec, Canada
\textsuperscript{b}Children's Hospital Boston, Boston, MA, USA

\textbf{A R T I C L E I N F O}

Article history:
Received 13 January 2010
Available online 15 October 2010

Keywords:
Influenza, Human
Population surveillance
Ambulatory care
Epidemiology
Syndrome
Medical Records Systems, Computerized
Seasons
Age factors
Immunization programs

\textbf{A B S T R A C T}

Syndromic surveillance is a novel automated approach to monitoring influenza activity, but there is no consensus regarding the most informative data sources for use within such a system. By comparing physician billing data from Quebec, Canada and hospital admission records, we assessed the timeliness of medical visits for influenza-like illnesses (ILI) to two types of outpatient healthcare settings. Overall, ILI visits by children aged 5–17 years at community-based settings were the most strongly correlated with hospital admissions and gave the greatest lead over hospital admissions. However, a degree of year-to-year variation suggests that syndromic surveillance of influenza should not focus on just a single subgroup. These findings reveal the richness of these real-time data for epidemic monitoring and demonstrate the flexibility of syndromic surveillance. By using real-time data, an evolving epidemic can be rapidly characterized by its epidemiological patterns, which is not possible with traditional surveillance systems.

\textcopyright 2010 Elsevier Inc. All rights reserved.

1. Introduction

Influenza is an infectious respiratory disease with an annual epidemic cycle associated with high mortality among the elderly [1,2], hospitalizations among both the elderly and the very young [1–3], and substantial economic consequences [4]. Recent epidemics of emerging infectious diseases such as SARS, avian influenza, and most recently, A/H1N1 “swine” influenza have stressed the need for better disease surveillance systems to enable prompt detection and aid the subsequent public health response.

Syndromic surveillance is a novel approach for monitoring disease occurrence. It relies on advanced statistical and computational methods to continuously monitor data that are collected automatically from clinical and other non-traditional settings in real-time or near real-time [5]. Since pre-diagnostic data are used, syndromic surveillance can provide an earlier indication of an outbreak than traditional laboratory or sentinel physician based surveillance [6]. Moreover, given the demographic data available from clinical sources, syndromic surveillance allows public health departments to monitor evolving patterns of illness among population subgroups, flexibly adjusting their focus as a situation emerges [7]. Several syndromic surveillance systems are described in the literature [8–11], and according to a survey, 89% of state, territorial and local public health departments in the United States perform syndromic surveillance [12]. Most systems follow data from emergency departments (ED) but there is little empirical evidence regarding the most informative data sources to monitor.

Although analyses of administrative data suggest that children are sentinels of an influenza epidemic in the population [2,13–15], there is debate concerning the specific age-groups that present earliest and provide the strongest indication of a rise in influenza activity. It is also not clear which setting provides the earliest signal as most studies have examined ED data [14–16] and few researchers have compared the timing of signals from visits to EDs to signals from visits to other settings such as community-based settings including private offices and community clinics.

Each influenza season can also be quite different due to the constant evolution of the influenza virus. The introduction of a new antigenic strain often leads to increased morbidity and healthcare utilization, as was the case with the B/Hong Kong/330/01 strain that spread to North America during the 2001–2002 season after a decade-long absence of the B/Victoria lineage on that continent.
2.2. Context

individual influenza epidemic period.

billing on a fee-for-service basis [18]. We suspected that annual differences in the healthcare utilization in a population may also modify the ability of a surveillance system to detect epidemic signals.

In this study, using physician billing data from Quebec, Canada, we sought to clarify the timing of visits for influenza-like illnesses (ILI) by age-group and setting, and to explore year-to-year variations in these patterns. Simultaneously, we also aimed to explore the utility of physician billing data from outpatient settings for automated syndromic surveillance of ILI.

2. Methods

2.1. Overview and study design

For our study period from January 4, 1998, to December 27, 2003 (inclusive), we compared the timing of age-restricted subsets of fee-for-service medical billing claims with ILI diagnoses from community-based care settings and hospital EDs (as a measure of ILI visits) to hospitalizations for P&I (as a robust measure of influenza circulation) in Quebec, Canada. In the first stage of our study, we examined overall patterns by including all weeks in our study period in a single analysis. In the second stage, we examined year-to-year variations by repeating separate analyses for each individual influenza epidemic period.

2.2. Context

The Régie de l’Assurance Maladie du Québec (RAMQ) is the agency responsible for the health insurance program in the province of Quebec. Ninety-nine percent of all Quebec residents are covered for the cost of hospital and physician services under this program, and of all medical visits billed to RAMQ, 85–95% are billed on a fee-for-service basis [18].

2.3. Data sources

2.3.1. Viral isolates data

Viral testing data were obtained from the Laboratoire de Santé Publique du Québec (Quebec Public Health Laboratory). These data included weekly counts of the results of three types of diagnostic tests for influenza (culture, antigen-detection, and polymerase chain reaction).

2.3.2. Fee-for-service billing claims data

In a previous study, we identified a set of 3424 new physicians who were licensed to practice in Ontario and/or Quebec, and then requested that RAMQ identify all patients seen by these physicians from 1998 through 2003 and provide us the fee-for-service billing claims submitted for these patients by any physician in Quebec (whether or not part of the set of study physicians) during this period [19]. Therefore, we had complete ascertainment of healthcare delivered on a fee-for-service basis for this cohort of patients. Each billing claim contains information such as anonymized unique identifiers for the physician and patient, an International Classification of Diseases, Ninth Revision (ICD-9) diagnostic code, a code for the setting type, and the date of the visit.

2.3.3. Hospitalization data

Our hospitalization data from the Quebec hospitalization database (MED-ECHO) captured all hospitalizations in Quebec. These records included the date of admission, date of discharge, and the discharge diagnosis.

2.4. Study population

We identified all physicians as “in practice” in the fee-for-service system by 1998 if they had at least one billing claim among our fee-for-service data from 1993 through 1997. The study population included all patients seen by these physicians from 1998 through 2003. Our study population in each year of our study period represented approximately 35–36% of the total source population of all RAMQ beneficiaries that had received at least one medical service in the same year [20]. Except for a slight overrepresentation of the elderly and females in our patient population, it was otherwise similar to the total RAMQ population by age and sex distributions.

2.5. Outcome measures

2.5.1. Epidemic period definition

Epidemic periods (see Table A-1 in Appendix A for details) were identified using viral isolates data, the gold standard for viral circulation. We pooled the results of the three diagnostic tests together and defined the start of each epidemic period as 4 weeks before the first two consecutive weeks during which the total number of positive specimens (for either influenza A or B) was five or more. We shifted the start week back to accommodate both our expectation that an increase in positive viral tests will be preceded by an increase in ILI visits, as well as the fact that we would later be shifting the time series in the cross-correlation function (CCF) computation. The end week was defined as the week before two consecutive weeks during which the total count was under five.

2.5.2. Medical visits for influenza-like illnesses (ILI)

A medical visit was considered an ILI visit if it resulted in at least one fee-for-service billing claim with an ICD-9 diagnostic code from the code set provided in Table A-2 in Appendix A. We constructed time series of the weekly number of ILI visits for each age-group and for each of two types of outpatient settings: (1) community-based care setting (including private offices, private clinics, and local community health and social services centers), and (2) hospital ED. We excluded hospital-based outpatient clinics from community-based care settings because they tend to see only referrals and cater to a different patient population compared to other community-based care settings in Quebec. During a single medical visit, a physician may perform multiple services, each of which would be billed for separately. To reduce overcounting of distinct visits, we aggregated billing claims such that each unique patient was counted no more than once per day in each time series.

2.5.3. Pneumonia and influenza (P&I) hospitalizations

We generated a time series of the number of short-term hospitalizations per week in all of Quebec with a primary discharge diagnosis of P&I (ICD-9 codes 480–487) to serve as a common reference against which the time series of ILI visits would be compared. P&I hospitalizations data are a commonly used measure for tracking and measuring the impact of influenza [21,22] because they provide a sensitive and representative measure of the burden of influenza morbidity [23,24].

2.6. Data analysis

Note: For a more detailed explanation of the statistical methods that were used, please see Appendix B (technical appendix).

2.6.1. Accounting for autocorrelation with ARIMA modeling

We first used Box and Jenkins seasonal autoregressive integrated moving average (ARIMA) models [25,26] to model the autocorrelation structure within each age-group and setting specific ILI
visits series. All models included two indicator variables to account for the effects of winter and non-winter holidays (listed in Table A-3 in Appendix A). For each time series, we assessed the adequacy of several candidate models using diagnostic criteria that test for the presence of autocorrelation in the residuals. The Akaike Information Criterion [26] was used to help choose between the models deemed adequate.

2.6.2. Analysis of timeliness and correlation through the cross-correlation function (CCF)

After finding ARIMA models for each subset of the ILI visits time series as described above, we applied the same model to the P&I hospitalizations reference time series. We then computed the cross-correlation function (CCF) at lags up to ±4 weeks between the residuals for each subpopulation's ILI visits time series and the residuals for the P&I hospitalizations time series (serving as a common reference series). In the first stage of our study, we produced one CCF using residuals for all weeks in our study period. However, in the second stage, we produced a separate CCF for each influenza epidemic period during our study period, and using only the residuals corresponding to epidemic weeks for each period. By applying ARIMA modeling and then using the residuals, we worked with a filtered time series in which the autocorrelation structure has been removed. Failure to account for temporal autocorrelation before correlating two time series can result in high correlations even when a true relationship between the variables does not exist [26–28]. To assess timeliness, we noted the lags in the CCF at which the maximum correlation, as well as the greatest lag at which a significant correlation were found for our overall analysis using data from all weeks of the study period. Table 3 shows selected subsets for the year-to-year analysis. See

3. Results

3.1. Descriptive statistics

From 1998 through 2003, there were a total of 2541,926 unique study patients with at least one billing claim from a community-based healthcare setting or a hospital ED in Quebec. Collectively, they made a total of 73,091,025 visits during the study period, 7% of which were given an ILI diagnosis (Fig. 1). There were higher proportions of females and young children among those diagnosed with ILI at least once during this period compared to the total study patient population (please see Table A-4 in Appendix A). During the same period, there were 104,571 short-term hospitalizations across Quebec with a primary diagnosis of P&I. The proportion of visits due to ILI was approximately the same (7%) among both community-based care settings and hospital EDs, although the majority of the ILI visits were to community settings (83%) as opposed to the ED (17%). A larger proportion of the ILI visits to community-based settings were from working-aged adults, while the elderly and the youngest children made up a greater share of ILI visits in emergency departments (Table 1).

3.2. Timeliness and correlation

For each cross-correlation function (CCF) between each subpopulation's ILI visits series and the P&I hospitalizations, Table 2 shows the lag at which the maximum correlation, as well as the greatest lag at which a significant correlation were found for our overall analysis using data from all weeks of the study period. Table 3 shows selected subsets for the year-to-year analysis. See

![Fig. 1. Time series plots of the weekly total counts of influenza-like illness (ILI) visits to community-based care settings and hospital emergency departments, and pneumonia and influenza (P&I) hospitalizations. Shaded regions indicate sustained periods of positive viral cultures (influenza A or B).](image-url)

For example, the heat-map of the CCFs for the overall analysis (Fig. 2) shows that visits by 5–12 year olds to community settings were most strongly correlated with P&I hospitalizations at the 2 week lag, but at the 3 week lag, ILI visits by 13–17 year olds to community settings showed the strongest standardized correlation with P&I hospitalizations.

3.2.2. Year-to-year analysis

The CCFs from the year-to-year analysis showed that correlations between outpatient ILI visits and P&I hospitalizations varied widely between different epidemic periods (Table A-6 in the appendix, and heat-maps in Fig. 3). The overall utility of ED visit data in terms of timeliness and strength of correlation are comparable to that of community settings in three seasons, better in one season (1998–1999) but worse in another (2001–2002). The settings and age-groups that were consistently correlated with P&I hospitalizations with the greatest lead time were community setting visits by 13–17 year olds and ED visits by 5–12 year olds (Table 3).

With subsets demonstrating at best a 2 week lead in most seasons, the 2001–2002 season was particularly distinctive for the 3 week lead for community setting visits by those aged <2, 2–4, and 13–17 years. The peak correlations for those aged 5–12, and 18–39 years occurred at the same lag but they fell just below statistical significance ($p = 0.05$). Even adults aged 40–64 years, unlike any other season, had a peak correlation at the 1 week lag.

4. Discussion

Using physician billing claims from outpatient care, we found that visits to community-based clinics ILI tended to increase in frequency earlier than visits to EDs. We also found that, relative to adults or the elderly, increases in the number of ILI visits by children provided the earliest signal of an epidemic. When considering both visit setting and age, community setting visits for ILI by children aged 2–17 years provided the greatest lead times over P&I hospitalizations. Overall, community setting visits by children aged 5–12 years stood out in particular due to the high peak correlation with P&I hospitalizations, although community-setting visits by children aged 13–17 years also exhibited a significant correlation at a greater lag compared to other subpopulations. These

### Table 1

<table>
<thead>
<tr>
<th>Age-group</th>
<th>Proportion of ILI visits to each setting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Community-based care settings (N = 4233782)</td>
</tr>
<tr>
<td>&lt;2 year</td>
<td>0.10</td>
</tr>
<tr>
<td>2–4 year</td>
<td>0.12</td>
</tr>
<tr>
<td>5–12 year</td>
<td>0.13</td>
</tr>
<tr>
<td>13–17 year</td>
<td>0.04</td>
</tr>
<tr>
<td>18–39 year</td>
<td>0.23</td>
</tr>
<tr>
<td>40–64 year</td>
<td>0.25</td>
</tr>
<tr>
<td>≥65 year</td>
<td>0.14</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Abbreviation: ILI, influenza-like illness.

### Table 2

<table>
<thead>
<tr>
<th>Subset</th>
<th>Age-group</th>
<th>Peak correlation</th>
<th>Earliest significant correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>All ages</td>
<td>Lag (weeks)</td>
<td>Correlation</td>
</tr>
<tr>
<td>Both setting types</td>
<td>All ages</td>
<td>0</td>
<td>0.29</td>
</tr>
<tr>
<td>By visit setting</td>
<td>Community-based settings&lt;sup&gt;+&lt;/sup&gt;</td>
<td>All ages</td>
<td>1</td>
</tr>
<tr>
<td>Emergency departments</td>
<td>All ages</td>
<td>0</td>
<td>0.45</td>
</tr>
<tr>
<td>By age-group</td>
<td>Community-based settings&lt;sup&gt;+&lt;/sup&gt;</td>
<td>&lt;2 year</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2–4 year</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5–12 year</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13–17 year</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18–39 year</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40–64 year</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>≥65 year</td>
<td>0</td>
<td>0.31</td>
</tr>
<tr>
<td>Emergency departments</td>
<td>&lt;2 year</td>
<td>0</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>2–4 year</td>
<td>0</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>5–12 year</td>
<td>2</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>13–17 year</td>
<td>2</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>18–39 year</td>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>40–64 year</td>
<td>0</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>≥65 year</td>
<td>0</td>
<td>0.57</td>
</tr>
</tbody>
</table>

<sup>1</sup> All correlations shown here were significant ($p = 0.05$).
<sup>2</sup> i.e. Private offices, private clinics and local community health and social services centers.
results have implications for implementing syndromic surveillance of ILI as they identify potential targets to monitor as sentinels of an oncoming influenza epidemic.

One reason for the earlier timing of ILI visits in community settings compared to the ED may be that the symptoms in the early stages of illness are mild enough that community-based care would be sought first over ED care [29]. However, the majority of those who have evaluated the potential utility of a particular data source for syndromic surveillance of influenza have focused on the ED [14–16, 30], and few have looked at community-based settings [31–34]. The predominance of ED-based studies is probably due to the fact that, especially in the United States, many of the already existing syndromic surveillance systems for influenza are based on ED data, such as chief complaints and discharge diagnoses, which are more easily obtainable compared to data from other settings. To our knowledge, only one study [13] has compared these two setting types, but data were drawn from several different populations. Using a single source population avoids potential confounding due to population differences in, for example, socioeconomic status or healthcare utilization behavior. Furthermore, that the province’s universal health insurance program covers 99% of all residents in our study location, and that this allowed us complete ascertainment of healthcare delivered on a fee-for-service basis for our patient study population presented us an opportunity to conduct a population-wide study that would not be easily possible in other locations such as the United States, and demonstrates the capacity of the Canadian universal health insurance programs for providing data to conduct such population-wide studies.

Our results for children are not surprising given that children have undeveloped immune systems that render them susceptible to influenza infection. They have the second highest rate of excess hospitalizations for P&I, after the elderly, and the highest rates of excess physician visits and ED visits for P&I (excess defined as beyond non-epidemic period baseline) [1].

Influenza morbidity is generally higher among younger children than for older children [2], but our results point to ILI visits from

Table 3
For each influenza epidemic period during 1998–2003, a cross-correlation function was computed between various age-group and setting-specific subsets of influenza-like illness visits, and a common reference time series of pneumonia and influenza hospitalizations in Quebec, Canada. This table shows selected subsets that demonstrated the greatest “lead times” based on (1) the peak correlation and (2) the earliest significant correlation (not necessarily peak) for each epidemic period. The subset that demonstrated the greatest correlation for each column and epidemic period is bolded.

<table>
<thead>
<tr>
<th>Subset Setting Age-group</th>
<th>Peak correlation</th>
<th>Earliest significant correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag (weeks)</td>
<td>Correlation</td>
</tr>
<tr>
<td>1998–1999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency department</td>
<td>2–4 year</td>
<td>2</td>
</tr>
<tr>
<td>Emergency department</td>
<td>5–12 year</td>
<td>3</td>
</tr>
<tr>
<td>Community-based setting</td>
<td>&lt;2 year</td>
<td>1</td>
</tr>
<tr>
<td>Community-based setting</td>
<td>13–17 year</td>
<td>2</td>
</tr>
<tr>
<td>Emergency department</td>
<td>18–39 year</td>
<td>0</td>
</tr>
<tr>
<td>1999–2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community-based setting</td>
<td>2–4 year</td>
<td>2</td>
</tr>
<tr>
<td>Community-based setting</td>
<td>5–12 year</td>
<td>2</td>
</tr>
<tr>
<td>Emergency department</td>
<td>2–4 year</td>
<td>0</td>
</tr>
<tr>
<td>Emergency department</td>
<td>5–12 year</td>
<td>0</td>
</tr>
<tr>
<td>2000–2001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community-based setting</td>
<td>&lt;2 year</td>
<td>3</td>
</tr>
<tr>
<td>Community-based setting</td>
<td>2–4 year</td>
<td>2,3</td>
</tr>
<tr>
<td>Community-based setting</td>
<td>13–17 year</td>
<td>3</td>
</tr>
<tr>
<td>2002–2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community-based setting</td>
<td>5–12 year</td>
<td>2</td>
</tr>
<tr>
<td>Community-based setting</td>
<td>13–17 year</td>
<td>2</td>
</tr>
<tr>
<td>Emergency department</td>
<td>5–12 year</td>
<td>2</td>
</tr>
<tr>
<td>Emergency department</td>
<td>13–17 year</td>
<td>2</td>
</tr>
</tbody>
</table>

* All correlations shown were significant (z = 0.05).

Correlation at lags of 2 and 3 weeks were the same.
school-aged children (5–17 years) as the earliest and strongest indicators of an influenza epidemic, which is consistent with their role as the primary vectors of influenza transmission in a population [35–38]. Their efficiency as spreaders of influenza may result from an interplay between both their innate ability to shed the virus earlier and for a longer time than adults [39,40], and their more extensive social contact patterns compared to younger infants [41,42]. In response to recent studies, for the 2008–2009 season, the Advisory Committee on Immunization Practices (ACIP) of the Centers for Disease Control and Prevention (CDC) in the United States updated their recommended target groups for annual vaccination to all children aged 6 months to 18 years [43]. Previously, among healthy children, the ACIP targeted only those aged 6 months to 4 years. Canada has not yet adopted this expanded vaccination policy [44].

Our findings, suggesting that pediatric age-groups act as a sentinel population for influenza within administrative data are consistent with other studies [2,13–15], although the exact age range is in dispute. Two other studies [2,14] identified school-aged children as sentinels, as we did, but others have pointed to slightly younger age-groups such as 3–4 years [13] and under 5 years [15].

However, we also found a degree of year-to-year variation in our results. Year-to-year variation may contribute to the reason why different pediatric age-groups have been identified as sentinel populations for influenza across different studies covering different study periods. The lack of consistently early rises in visits in specific age-groups and settings has important implications for setting and age-group focused influenza surveillance. This finding also suggests that it will be difficult to construct accurate age-specific statistical forecasting models because such models require stable indicators of the onset or peak of an outbreak [45].

This year-to-year variation in the age-groups and settings with the earliest rise in visits may be a consequence of the constant evolution of the influenza virus that results in the regular emergence of new strains. If the mutation rate is fast, herd immunity is decreased as fewer individuals will have had the opportunity to gain immunity to circulating strains through exposure. A/H3N2 viruses are believed to have faster rates of antigenic mutation (antigenic drift) than A/H1N1 and B viruses [46]. It is for this reason that years predominated by A/H3N2 strains are associated with more severe epidemics, especially for young children and the elderly [47–49].

During the 1998–1999 season, an A/H3N2 strain was the predominant strain in circulation, and we found that ILI visits to the ED were significantly correlated with P&I hospitalizations at lags of 1 week or more for all age-groups except those <2 years. On the other hand, no community setting subset provided any lead during this season. In contrast, for the 2000–2001 season, during which no laboratory-confirmed cases of subtype A/H3N2 were reported in Quebec [50], no ED subset provided any lead. Since the ED typically sees more severe cases than community settings, the contrast in lead times may reflect heavier ED utilization during the more severe A/H3N2 seasons.

A relationship between age and influenza subtype has been established as well. Fox et al. found that young school children (5–9 years) had the highest infection rates and were the main introducers of influenza during A/H3N2 seasons, but implicated teenagers (10–19 years) during A/H1N1 and B seasons [35,51]. Other study pointed to younger age-groups: those aged 1–4 years during type A outbreaks and those aged 5–9 years during mixed or type B outbreaks [52]. In our study, community visits by 13–17 year olds provided the earliest leads in two of the three A/H3N2 predominant seasons (1999–2000, 2001–2002). In contrast, community visits by 5–12 year olds provided the best lead during the single A/H1N1 and B predominant season (2000–2001) in our study period. While these age–strain interactions are worthy of further investigation, the variation in the findings across these different studies and the small number of influenza seasons examined means definitive conclusions about the dependence of timeliness on age and influenza subtype cannot be made.

Reflecting the link between age and immunity, when a strain re-emerges after a long absence, children will often be particularly vulnerable to infection, unlike adults who may already have immunity if the strain last circulated within their life time. Furthermore, a vaccine mismatch usually occurs for these seasons. The 2001–2002 season, when a B/Victoria-lineage virus re-emerged for the first time in a decade in North America [53], was particularly severe, especially for school-aged children, and ED visits [14]. Similarly, the emergence of a novel A/H1N1 influenza strain in 2009 disproportionately affected children and young adults, with many school-based outbreaks [54–57], as is characteristic of pandemic influenza [58]. Our results for the 2001–2002 season demonstrate the impact of a re-emerged strain on the timing of ILI visits by...
children to community settings as well. For our study location (Quebec), the B/Victoria lineage had last been identified during the 1988–1989 season [59]. Therefore it would be expected that most of the younger children (<13 years) had no or limited immunity. Among community setting subsets, we found a remarkable 3 week lead for most of the pediatric age-groups during this season. However, no lead was observed for the two oldest age-groups (40–64 and ≥65 years), consistent with their presumed prior immunity to this lineage.

Although a year-by-year analysis may not always be helpful in deciding which subgroup may be most likely to provide the strongest and earliest signals for influenza surveillance each year, this approach may nonetheless be a beneficial complement to an overall analysis, as the yearly variation observed in this and other studies suggests [15,16,60]. Analyzing yearly variations may prevent inappropriate generalizations or reveal patterns underlying different years with common traits.

There are some limitations to our study. Respiratory syncytial virus (RSV) is another major viral respiratory pathogen whose impact exceeds that of influenza among young children [61] and the elderly [62]. We could not distinguish RSV from influenza since RSV is clinically similar to and often co-circulates with influenza [63]. More broadly, misclassification bias (misdiagnosis or errors in coding) is a chief limitation in the use of ICD-9 coded billing data compared to laboratory confirmed diagnoses. The patterns in our time series ofILI visits may reflect a combination of the patterns of both influenza and RSV, which could diminish the correlation between the ILI visits time series and the P&I hospitalizations time series. However, RSV mainly affects the very young, and we used an ILI code set that has been validated against influenza viral isolates [31]. We also acknowledge that our epidemic period definition is unverified, but there has been no consistent definition and a variety of approaches have been used [1,15,60]. We tried other slight variations in the epidemic period definition with little impact on the results. With only five influenza seasons’ worth of data, it is also difficult to make generalizations for specific influenza subtypes. Finally, while we argue for the utility of these real-time or near real-time data for surveillance, we should also note that it may be difficult to get similar data in real-time or near real-time in the United States and other locations. While we have looked at syndromic data only in this study, it would also be important to also conduct a separate comparison of the value of traditional “pen and paper” sentinel clinic site surveillance to that of electronic automated syndromic surveillance for ILI to validate the advantages of one over the other.

5. Conclusion

Across 6 years, medical visits by school-aged children to community clinics provided the earliest indication of the onset of an influenza epidemic. Separate analysis of each influenza season revealed that ILI syndromic surveillance should not focus on any single subgroup but a combination of several age-group and visit setting specific subgroups. Due to region-specific differences in healthcare seeking behaviors, we recommend that our findings be replicated for different regions. This study demonstrates the flexibility of automated syndromic surveillance for monitoring real-time data as it allows for the rapid identification of age and setting epidemiological patterns in the data. This kind of insight is important for both seasonal influenza and for novel and rapidly evolving outbreaks that emerge unexpectedly and for which the epidemiological profile is unknown. A better understanding of these relationships would help improve the accuracy of infectious disease surveillance systems, and enable the prompt initiation and evaluation of appropriate public health interventions.

Acknowledgments

We thank Jean Gratton at the Direction de Santé Publique de Montréal and Rodica Gilca at the Institut National de Santé Publique du Québec for their assistance with aspects of the data, and members of our research group for their comments on the manuscript. This research was supported by a scholarship from the McGill Center for Bioinformatics (Emily Chan), a Canada Research Chair in Public Health Informatics (David Buckner), and an operating grant (PAN–83152) from the Canadian Institutes of Health Research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jbi.2010.10.001.

References


