Perspective

Characteristics of US public schools with reported cases of novel influenza A (H1N1)

Anne Gatewood Hoen a,b,c, David L. Buckeridge c,d, Emily H. Chan a, Clark C. Freifeld a,e, Mikaela Keller a,b, Katia Charland a,b,c, Christl A. Donnelly f, John S. Brownstein a,b,*

a Children’s Hospital Informatics Program, Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts, USA
b Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA
c Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
d Direction de sante´ publique de Montre´al, Que´bec, Canada
e MIT Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
f MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, London, UK

1. Introduction

One striking feature of the 2009 pandemic of novel influenza A (H1N1) has been the skewed age distribution of confirmed cases, with children and young adults disproportionately affected.1 Early clinical evidence revealed a shift in the age distribution, with the majority of deaths and severe cases being patients between the ages of 5 and 59 years, and especially children.2–4 Possible explanations for this pattern of infection include preexisting immunity in older age groups and an important role for schools as settings for the early outbreaks of the pandemic.1,5

We monitored in real-time the early school-related outbreaks of novel H1N1 influenza in the USA using the event-based HealthMap disease surveillance platform.6–8 HealthMap monitors informal web-based media reports, moderated distribution lists such as ProMED Mail, and official public health agency alerts for disease outbreak information. We analyzed early media reports collected by HealthMap for information on novel H1N1 influenza outbreaks in schools in order to build an early epidemiological picture of the novel H1N1 epidemic in US public schools.9

2. Methods

Between April 23 and June 8, 2009, HealthMap detected 181 English-language media reports related to suspected or confirmed cases of novel H1N1 influenza in schools and universities worldwide. We filtered these reports to examine more closely public primary and secondary schools in the USA with one or more confirmed cases of H1N1 influenza and 6815 control schools located in the same 23 counties as case schools.

Results: Compared with controls from the same county, schools with reports of confirmed cases of H1N1 influenza were less likely to have a high proportion of economically disadvantaged students (adjusted odds ratio (aOR) 0.385; 95% confidence interval (CI) 0.166–0.894) and less likely to have older students (aOR 0.792; 95% CI 0.670–0.938).

Conclusions: We conclude that public schools with younger, more affluent students may be considered sentinels of the epidemic and may have played a role in its initial spread.

© 2010 International Society for Infectious Diseases. Published by Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +1 617 355 6998; fax: +1 617 730 0267.
E-mail address: john.brownstein@childrens.harvard.edu (J.S. Brownstein).
We used a backward elimination model selection procedure to build a multivariate model to estimate the probability of a school having one or more confirmed cases detected by HealthMap. The number of students enrolled in the school was log-transformed and included in all models to account for the probability of one or more confirmed cases as a function of student body size. Because schools with media reports were matched to control schools within the same county, we used conditional logistic regression to account for within-county dependence. The R Statistical System (version 2.7.2, R Foundation for Statistical Computing, http://www.R-project.org) was used for all statistical computations.

### 3. Results

Relationships between school characteristics and media reports of novel H1N1 are presented in Table 1. The final multivariate model revealed independent significant relationships with number of students (adjusted odds ratio (aOR) 7.344, 95% confidence interval (CI) 3.100–17.398; \( p < 0.001 \)), highest grade level (aOR 0.792, 95% CI 0.670–0.938; \( p = 0.001 \)), and Title 1 status (aOR 0.385, 95% CI 0.166–0.894; \( p < 0.001 \)). As expected, schools with more students were more likely to have been reported as having one or more confirmed cases of novel H1N1 influenza. In addition, schools with lower maximum grade levels (in general, primary schools) and schools not qualifying for Title 1 funding (schools with fewer economically disadvantaged students) were more likely than other schools in the same county to have been detected. Lowest grade level, grade span, student-to-teacher ratio, the urbanized area-indicator, and the variables relating to racial/ethnic makeup of schools were dropped from the final multivariate model. Overall, this analysis suggests that within affected counties, affluent schools with a younger student body are more likely than other schools in the same community to have confirmed cases of novel H1N1 influenza that are picked up by the media and detected by HealthMap.

### 4. Discussion

We have presented an initial characterization of the US public schools affected by the recent novel H1N1 influenza outbreak using a real-time, informal surveillance system, HealthMap. While there is no tool for monitoring outbreaks of this nature in the US that is without detection biases, there are some limitations specific to our approach. Namely, we were limited not only by the ability of public health officials to detect and confirm cases, but also by the

![Counts of cases](image)

**Figure 1.** Map of counties showing schools with one or more confirmed cases of H1N1 influenza detected by HealthMap.

### Table 1

School characteristics and conditional logistic regression models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean ± SD or n (%)</th>
<th>Other schools in county (n=6815)</th>
<th>Univariate/bivariate models’ OR (95% CI)</th>
<th>p-Value</th>
<th>Final multivariate model aOR (95% CI)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student body size</td>
<td>945.22 ± 677.21</td>
<td>704.10 ± 618.54</td>
<td>3.759 (2.158, 6.457)</td>
<td>&lt;0.001</td>
<td>7.344 (3.100, 17.398)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lowest grade at school</td>
<td>3.38 ± 4.19</td>
<td>2.42 ± 3.81</td>
<td>0.929 (0.832, 1.036)</td>
<td>0.185</td>
<td>0.792 (0.670, 0.938)</td>
<td>0.007</td>
</tr>
<tr>
<td>Highest grade at school</td>
<td>7.53 ± 3.20</td>
<td>7.16 ± 2.93</td>
<td>0.828 (0.703, 0.975)</td>
<td>0.024</td>
<td>0.991 (0.872, 1.126)</td>
<td>0.893</td>
</tr>
<tr>
<td>Grade span</td>
<td>4.16 ± 2.38</td>
<td>4.74 ± 2.33</td>
<td>0.961 (0.881, 1.153)</td>
<td>0.671</td>
<td>1.607 (1.340, 1.943)</td>
<td>0.118</td>
</tr>
<tr>
<td>Student–teacher ratio</td>
<td>16.08 ± 3.45</td>
<td>17.28 ± 4.98</td>
<td>0.991 (0.872, 1.126)</td>
<td>0.185</td>
<td>0.991 (0.872, 1.126)</td>
<td>0.185</td>
</tr>
<tr>
<td>Title 1 school (%)</td>
<td>14 (43.8)</td>
<td>4694 (68.9)</td>
<td>0.525 (0.234, 1.178)</td>
<td>0.118</td>
<td>0.607 (0.340, 1.063)</td>
<td>0.118</td>
</tr>
<tr>
<td>Located in urbanized area (%)</td>
<td>24 (75.0)</td>
<td>6311 (92.6)</td>
<td>0.525 (0.234, 1.178)</td>
<td>0.118</td>
<td>0.607 (0.340, 1.063)</td>
<td>0.118</td>
</tr>
<tr>
<td>Proportion Hispanic students</td>
<td>0.13 ± 0.21</td>
<td>0.37 ± 0.32</td>
<td>0.180 (0.013, 2.467)</td>
<td>0.199</td>
<td>0.180 (0.013, 2.467)</td>
<td>0.199</td>
</tr>
<tr>
<td>Proportion white students</td>
<td>0.55 ± 0.35</td>
<td>0.31 ± 0.32</td>
<td>4.611 (0.627, 33.894)</td>
<td>0.133</td>
<td>4.611 (0.627, 33.894)</td>
<td>0.133</td>
</tr>
<tr>
<td>Proportion black students</td>
<td>0.09 ± 0.19</td>
<td>0.10 ± 0.27</td>
<td>0.154 (0.008, 2.897)</td>
<td>0.211</td>
<td>0.154 (0.008, 2.897)</td>
<td>0.211</td>
</tr>
<tr>
<td>Proportion Asian students</td>
<td>0.23 ± 0.32</td>
<td>0.12 ± 0.18</td>
<td>0.939 (0.042, 21.204)</td>
<td>0.533</td>
<td>0.939 (0.042, 21.204)</td>
<td>0.533</td>
</tr>
<tr>
<td>Racial/ethnic diversity</td>
<td>0.32 ± 0.20</td>
<td>0.40 ± 0.21</td>
<td>0.450 (0.033, 6.000)</td>
<td>0.546</td>
<td>0.450 (0.033, 6.000)</td>
<td>0.546</td>
</tr>
</tbody>
</table>

SD, standard deviation; OR, odds ratio; aOR, adjusted odds ratio; CI, confidence interval.

Student body size model is univariate. All others are bivariate, with log-transformed student body size as the unlisted explanatory variable.
ability and willingness of the media to report them. This was less of an issue in the early stages of the epidemic, when both sectors were actively investigating outbreaks, and for this reason we limited our analysis to the earliest news reports. Detailed evaluation of the utility of these data sources remains an important research question. However, our approach allowed us to quickly detect and catalog, in real-time, outbreaks of novel H1N1 influenza in schools that, to our knowledge, were not otherwise formally documented at this scale. Influenza transmission in schools is not reportable to the state or federal governments, making the news media a potentially more sensitive and timely data source than the current voluntary systems in place.

We compared schools that were reported to have experienced outbreaks of novel H1N1 influenza with nearby schools that were not, and we have presented an initial investigation of school characteristics associated with such reporting. Our observation that schools with lower grade levels were likely to have an early case report detected by HealthMap is consistent with previous reports implicating the younger pediatric age groups as early sentinels of seasonal influenza epidemics. However, it is not entirely clear why more affluent schools with lower grade levels were more likely than other schools in the same county to report cases. Given the relatively mild clinical outcomes of the first cases, it could be that students attending these schools were more likely than others to have their illness diagnosed by a physician and thus reported as confirmed cases, or that these schools were more likely to receive media coverage when a case was detected. Because many of the first school-related cases of the epidemic were the result of students with travel histories to Mexico, it is also possible that these schools were more likely to have students who had recently traveled to Mexico. It is unclear whether the characteristics of schools identified represent actual risk factors for infection, or whether they characterize schools that may be most sensitive to case detection and media reporting. Nonetheless, the novel influenza A (H1N1) influenza pandemic has illustrated the importance of understanding the best targets for surveillance efforts to allow for early, sensitive, and accurate outbreak detection. Informal, event-based disease surveillance tools such as HealthMap and others hold promise as useful technologies that enable real-time assessments of epidemiological characteristics prior to the availability of such information through more traditional surveillance methods.

Acknowledgements

We gratefully acknowledge Amy Sonricker, Susan Aman, Rebecca F. Baggaley, and T. Déirdre Hollingsworth for helpful discussions and computational support.

Funding support: This work was supported by R21AI073591-01 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, PAN-83152 and CAT-86857 from the Canadian Institutes of Health Research, and a research grant from Google.org. CAD acknowledges the Medical Research Council, UK, for funding support. Neither the funding sources nor the sponsor played a role in the design and conduct of the study; the collection, management, analysis, and the interpretation of the data; or in the preparation, review, and approval of the manuscript.

Conflict of interest: The authors have no conflicts of interest to report.

References