Predicting virologically-confirmed influenza using school absences during the 2007-2015 seasons in Allegheny County, PA
Talia M. Quandelacy1, Shanta Zimmer2, Chuck Vukotich2, Rachael Bieltz2, Kyra Granitz2, David Galloway2, Justin Lessler1, Yenlik Zheteyeva2, Hongjiang Gao3, Amra Uzicanin4, Derek A.T. Cummings1,3
1. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
2. School of Medicine, University of Pittsburgh, Pittsburgh, PA
3. Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA
4. Department of Global Migration and Quarantine, Centers for Disease Control and Prevention, Atlanta, GA

BACKGROUND
• School-based surveillance has been considered for real-time flu monitoring, as 5-17 year olds play an important role in community-level transmission.
• We studied if all-cause and cause-specific school absences improved predictions of virologically-confirmed influenza in the community.

RESULTS
• Overall, in Allegheny county, there were:
  - 2,184,220 reported all-cause absences from 9 school districts (2010-2015)
  - 19,577 reported all-cause and 3,012 ILI-related absences from 3 school districts (2007, 2012 and 2015)
• Including 1-wk lagged absences rates in multivariate models improved model fits & predictions of ILI cases over models with weekly and weekly average temperature (ΔAIC ≈ -4).
• All-cause absences from lower grades explained data best.
  - Kindergarten absences explained 22.1% of model deviance compared to 0.43% using 12 grade absences in validation.
• Multivariate models of week-lagged kindergarten absences, week of year, & weekly average temperature had the best fits over other grade-specific multivariate models (ΔAIC ≈ 6 comparing K to 12th grade).
• ILI-specific absences performed marginally better than total absences in two years, adjusting for other covariates, but markedly worse in one year. However, these results were based on a small number of observations.
• Increased length of absence from school also improved predictions.

DATA & METHODS
• Virologically-confirmed flu cases (all ages) collected from all county emergency departments & outpatient providers (2007 and 2010-2016) provided by Allegheny County Department of Health.
• Reported all-cause school absence rates from 9 Allegheny County school districts for 2010-2015.
• Six school districts provided influenza-like illness (ILI)-specific absences collected using a standard protocol:
  - 10 K-5 schools in 1 school district (2007-2008)
  - 9 K-12 schools in 2 school districts (2012-2013)
  - 9 K-12 schools from 3 school districts (2015-2016)
• We used negative binomial regressions to predict weekly county-level flu cases in Allegheny County, Pennsylvania during the 2010-2015 seasons.
• Candidate model covariates:
  - All-cause school absence rates of different weekly (1-3) log & administrative levels (county, school type, & grade) (assessed separately)
  - Week and month of the year (assessed separately)
  - Average weekly temperature & relative humidity (assessed separately)
• Separately for 3 districts for which ILI-specific and all-cause absences were available, we predicted weekly county-level influenza cases using all-cause and ILI-specific absences with all previously stated covariates.
• We used several cross-validation approaches to assess models including leave 20% of weeks out, leave 20% of schools out, and leave 52-weeks out.

CONCLUSIONS
• Our findings suggest models including younger student absences improve predictions of virologically-confirmed influenza.
• We found ILI-specific absences performed similarly to all-cause absences, but more observations are needed to assess the relative performances of these two datasets.

REFERENCES
This study is supported under the OMS grant # 1 U01 CI 0000775-01.
1. CDC, supporting the Johns Hopkins Center for Disease Control and Prevention, and the Department of Epidemiology at Johns Hopkins School of Public Health. Student travel to this conference is supported by the MD/USF SPS program.

Table 3. All-cause and cause-specific model performance using three school-based cohort data for three influenza seasons

<table>
<thead>
<tr>
<th>Influenza Season/Coast</th>
<th>Absence-type model</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007-2008</td>
<td>All-cause absence</td>
<td>44.3%</td>
</tr>
<tr>
<td>2010-2013</td>
<td>SMART1</td>
<td>90.9%</td>
</tr>
<tr>
<td>2015-2016</td>
<td>SMART2</td>
<td>90.0%</td>
</tr>
</tbody>
</table>

SMART1 includes weekly lagged absence rates from the previous week, week of the year, average temperature, SMART2 modeled includes weekly lagged absence rates and week of the year (Pearson correlation 0.34), leave 20% of school out, R² was obtained by linearly regressing the observed cases against the predicted cases using 100% of sample data.

Table 2. Performance of models including week-lagged kindergarten absences to predict virologically-confirmed influenza in Allegheny County, PA.

<table>
<thead>
<tr>
<th>Validation</th>
<th>Model</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
<th>Leave 20% of data out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influenza Season/Coast</td>
<td>Absence-type model</td>
<td>R²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td>All-cause absence</td>
<td>44.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2013</td>
<td>SMART1</td>
<td>90.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-2016</td>
<td>SMART2</td>
<td>90.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Smart1 includes weekly lagged absence rates from the previous week, week of the year, average temperature, SMART2 modeled includes weekly lagged absence rates and week of the year (Pearson correlation 0.34), leave 20% of school out, R² was obtained by linearly regressing the observed cases against the predicted cases using 100% of sample data.

FIGURE 1. ALL-CAUSE ABSENCE, ILI-SPECIFIC ABSENCE & REPORTED CONFIRMED FLU CASES OVER FIVE INFLUENZA SEASONS

FIGURE 2. OBSERVED & PREDICTED CASES USING WEEKLY COUNTY ALL-CAUSE ABSENCE

FIGURE 3. GRADE SPECIFIC MODEL DEVIANE AND PREDICTIONS USING KINDERGARTEN ABSENCE

FIGURE 4. OBSERVED & PREDICTED FLU USING ALL-CAUSE VS ILI-SPECIFIC ABSENCE

Table 1. All-cause and cause-specific model performance using three school-based cohort data for three influenza seasons

<table>
<thead>
<tr>
<th>Influenza Season/Coast</th>
<th>Absence-type model</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007-2008</td>
<td>All-cause absence</td>
<td>44.3%</td>
</tr>
<tr>
<td>2010-2013</td>
<td>SMART1</td>
<td>90.9%</td>
</tr>
<tr>
<td>2015-2016</td>
<td>SMART2</td>
<td>90.0%</td>
</tr>
</tbody>
</table>

SMART1 includes weekly lagged absence rates from the previous week, week of the year, average temperature, SMART2 modeled includes weekly lagged absence rates and week of the year (Pearson correlation 0.34), leave 20% of school out, R² was obtained by linearly regressing the observed cases against the predicted cases using 100% of sample data.

Modern values are calculated using delta weekly lagged absences, and average values of the week and average temperature. SMART1 modeled includes weekly lagged absence rates from the previous week, week of the year, average temperature, SMART2 modeled includes weekly lagged absence rates and week of the year (Pearson correlation 0.34), leave 20% of school out, R² was obtained by linearly regressing the observed cases against the predicted cases using 100% of sample data.

CONCLUSIONS
• Our findings suggest models including younger student absences improve predictions of virologically-confirmed influenza.
• We found ILI-specific absences performed similarly to all-cause absences, but more observations are needed to assess the relative performances of these two datasets.

REFERENCES
This study is supported under the OMS grant # 1 U01 CI 0000775-01.
1. CDC, supporting the Johns Hopkins Center for Disease Control and Prevention, and the Department of Epidemiology at Johns Hopkins School of Public Health. Student travel to this conference is supported by the MD/USF SPS program.