Predicting virologically-confirmed influenza using school absences during the 2007-2015 seasons in Allegheny County, PA

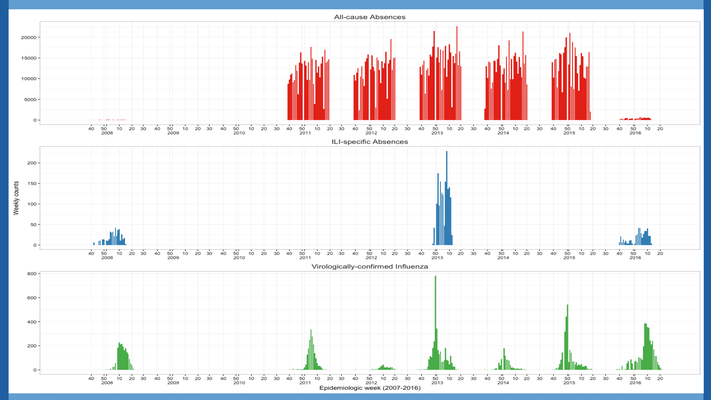
Talia M. Quandelacy[1], Shanta Zimmer[2], Chuck Vukotich[2], Rachael Bieltz[2], Kyra Grantz[3], David Galloway[2], Justin Lessler[1], Yenlik Zheteyeva[4], Hongjiang Gao[4], Amra Uzicanin[4], Derek A.T. Cummings[1,3]

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 Biology and Emerging Pathogens Institute, University of Florida, Gainesville, FL
4. Division 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.

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



Weekly counts of reported all-cause absences (9 school districts (SD), 2010-2015, a subset of schools in 3 districts 2007-2008, 2015-2016) (top panel). ILI-specific absences for 3 SDs (2007-2008, 2012-2013 and 2015-2016) (middle panel) and virologically-confirmed influenza all of Allegheny county from 2007 and 2010-2016 (bottom panel).

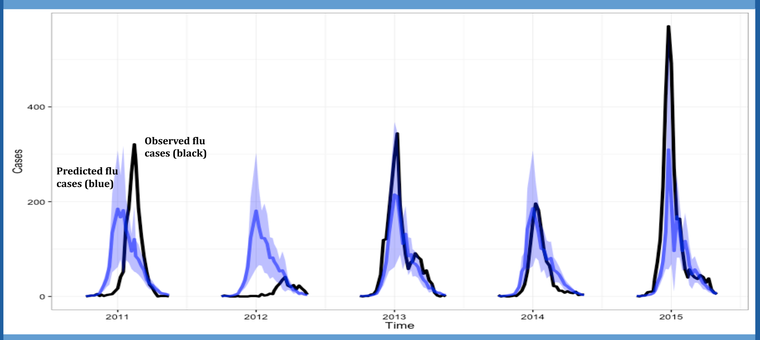
DATA & METHODS

- 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) lags & 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.

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)
  - 11,946 reported virologically-confirmed influenza cases (2007, 2010-2015)
- Including 1-wk lagged absence rates in multivariate models improved model fits & predictions of influenza cases over models using week of year 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 12th 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 perform 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

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



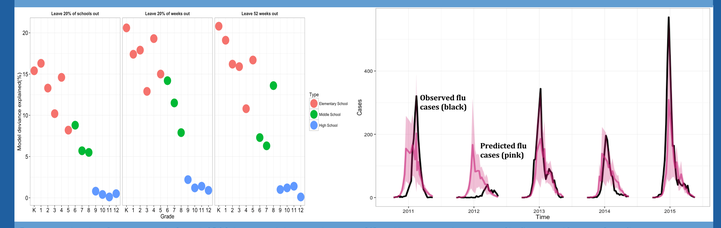
Weekly observed virologically-confirmed (black line) and predicted (blue line) influenza cases over five influenza seasons (2010-2015) in Allegheny County . Predictions used a negative binomial regression accounting for 1-week lagged county absences, week of the year, and average temperature. Blue shading represents the uncertainty bounds. Predictions shown here use 20% of schools left out validation.

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

Validation	Model	R^2
Leave 20% of data out	Absence	58.1%
Leave 20% of data out	Temperature & week	35.9%
Leave 20% of data out	Absence, temperature & week	65.4%
Leave 52 weeks out	Absence	19.2%
Leave 52 weeks out	Temperature & week	49.9%
Leave 52 weeks out	Absence, temperature & week	42.8%
Leave 20% of schools out	Absence	19.1%
Leave 20% of schools out	Temperature & week	40.6%
Leave 20% of schools out	Absence, temperature & week	58.6%

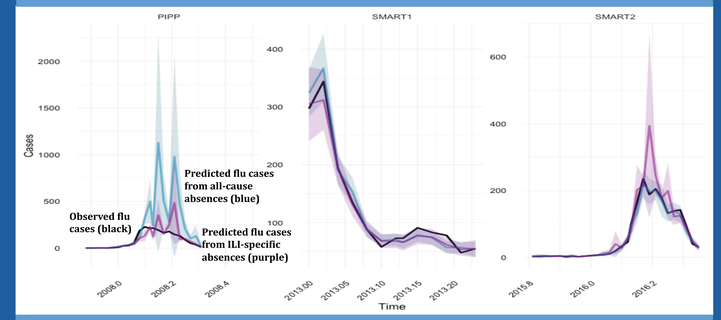
Negative binomial model included various lagged weekly absence rates, and splines of week of the year, and average temperature. Model validation withheld 20% of data. Deviance is the in-sample deviance explained by the model. R^2 was obtained by linearly regressing the observed cases against the predicted cases using the out-of-sample data. Mean absolute error is the mean absolute prediction error over the time-series.

FIGURE 3. GRADE-SPECIFIC MODEL DEVIANCE AND PREDICTIONS USING KINDERGARTEN ABSENCES



In-sample deviance explained (%) by weekly grade-specific all-cause absences (left panel) ranging from kindergarten to 12th grade, & observed & predicted cases from kindergarten absences (right panel) using negative binomial regression models of virologically-confirmed influenza in Allegheny County from 2010-2015. Colors in left panel represent different school types: red for elementary school grades, green for middle school grades, and blue for high school grades. Pink line & shading represent the predicted cases and 95% uncertainty bounds using a negative binomial regression of absences, wk & temp, & leave 20% of schools out validation.

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



Weekly virologically-confirmed observed (black) and predicted (blue line refers to all-cause absence model and purple line refers to ILI-specific absence model) influenza cases during 3 seasons (2007-2008, 2012-2013, & 2015-2016). Shading represents uncertainty bounds. model included weekly-lagged absence rates from the previous week, week of the year, and average temperature. SMART[1] models included weekly-lagged absence rates and week of the year. Cross-validation used leave 20% of schools out.

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

Influenza Season	Cohort	Absence-type model	R^2
2007-2008	PIPP	All-cause absence	44.3%
		ILI-specific absence	49.0%
2012-2013	SMART[1]	All-cause absence	99.9%
		ILI-specific absence	99.9%
2015-2016	SMART[2]	All-cause absence	99.0%
		ILI-specific absence	84.0%
Pooled analysis	PIPP, SMART[1], SMART[2]	All-cause absence	35.0%
		ILI-specific absence	31.0%

Each negative binomial model included weekly-lagged absence rates from the previous week, week of the year, and average temperature. SMART[1] models included weekly-lagged absence rates and week of the year. Cross-validation used leave 20% of schools out. R^2 was obtained by linearly regressing the observed cases against the predicted cases using the out-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.

Acknowledgements
This study is supported under the CDC grant 1 U01 CK000337-01.
TMQ is supported by the Johns Hopkins Center for American Indian Health, and the Department of Epidemiology Thesis Research Fund. Student travel to this conference is supported by the NSF/AISES LTP program