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

Talia Quandelacy[1], Shanta Zimmer[2], Justin Lessler[3], Charles VUKOTICH[4], Rachel Bieltz[4], Kyra Grantz[5], David Galloway[4], Jonathan Read[6], Yenlik Zheteyeva[7], Hongjiang Gao[7], Amra Uzicanin[8], and Derek Cummings[5]

[1]Johns Hopkins University Bloomberg School of Public Health
[2]University of Pittsburgh School of Medicine
[3]Johns Hopkins Bloomberg School of Public Health
[4]University of Pittsburgh
[5]University of Florida
[6]University of Liverpool
[7]Centers for Disease Control and Prevention
[8]Affiliation not available

February 14, 2021

Abstract

Background Children are important in community-level influenza transmission. School-based monitoring may inform influenza surveillance. Methods We used reported weekly confirmed influenza in Allegheny County during the 2007, and 2010-2015 influenza seasons using Pennsylvania's Allegheny County Health Department all-age influenza cases from health facilities, and all-cause and influenza-like illness (ILI)-specific absences from nine county school districts. Negative binomial regression predicted influenza cases using all-cause and illness-specific absence rates, calendar week, average weekly temperature and relative humidity, using four cross-validations. Results School districts reported 2,184,220 all-cause absences (2010-2015). Three one-season studies reported 19,577 all-cause and 3,012 ILI-related absences (2007, 2012, 2015). Over seven seasons, 11,946 confirmed influenza cases were reported. Absences improved seasonal model fits and predictions. Multivariate models using elementary school absences outperformed middle and high school models (relative mean absolute error (relMAE)=0.94, 0.98, 0.99). K-5 grade-specific absence models had lowest mean absolute errors (MAE) in cross-validations. ILI-specific absences performed marginally better than all-cause absences in two years, adjusting for other covariates, but markedly worse one year. Conclusions Our findings suggest seasonal models including K-5th grade absences predict all-age confirmed influenza and may serve as a useful surveillance tool.

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

Running Title: School absences predicting county influenza

Talia M. Quandelacy[1,2], Shanta Zimmer[3,4], Justin Lessler[1], Charles Vukotich[3], Rachel Bieltz[3], Kyra H. Grantz[5], David Galloway[3], Jonathan M. Read[6], Yenlik Zheteyeva[7], Hongjiang Gao[7], Amra Uzicanin[7], Derek A.T. Cummings[1,5]

Affiliations:

1. Johns Hopkins University, Baltimore, MD, USA (TMQ taliaquandelacy@jhu.edu, JL justin@jhu.edu, DATC datc@ufl.edu)

2. University of Colorado, Anschutz Medical Campus, Aurora, CO (TMQ talia.quandelacy@cuanschutz.edu)

3. University of Pittsburgh, Pittsburgh, PA, USA (CV charlesv@pitt.edu, RB rlb105@pitt.edu, DG ddg5@pitt.edu)

4. University of Colorado, Denver, CO, USA (SMZ shanta.zimmer@ucdenver.edu)

5. University of Florida, Gainesville, FL, USA (KHG kgrantz@ufl.edu, DATC datc@ufl.edu)

6. Lancaster University, Lancaster, United Kingdom (JMR jonathan.read@lancaster.ac.uk)

7. Centers for Disease Control and Prevention, Atlanta, GA, USA (YZ igg0@cdc.gov, HG uxi7@cdc.gov, AU aau5@cdc.gov)

Corresponding Author: Talia Quandelacy (talia.quandelacy@cuanschutz.edu)

Conflict of interest

The authors declare that they have no conflicts

Authorship

DATC, and JL contributed to concept and design of analysis. TMQ contributed to analyses. TMQ, DATC, JMR, and JL drafted and revised manuscript. TMQ, DATC, JL, CV, SZ, KG, RB, DG, YZ, BD, HG, and AU gave final approval.

Acknowledgments

This work was supported by the US Centers for Disease Control and Prevention (Cooperative Agreement 1 U01 CK000337-01 to DATC), and the Engineering and Physical Sciences Research Council (grant EP/N014499/1 to JMR). We would like to thank the Allegheny County Department of Health (Dr. Luanne Brink and Steve Forest) for providing their data; and the schools, staff, students, and parents who participated in PIPP, SMART, and SMART^2. We also thank Rahaan Gangat of the National Weather Service Pittsburgh for his assistance with climate data. We thank Scott Zeger for helpful discussions regarding analyses.

Word count: 3,204

Abstract word count: 184

Abstract

Background

Children are important in community-level influenza transmission. School-based monitoring may inform influenza surveillance.

Methods

We used reported weekly confirmed influenza in Allegheny County during the 2007, and 2010-2015 influenza seasons using Pennsylvania's Allegheny County Health Department all-age influenza cases from health facilities, and all-cause and influenza-like illness (ILI)-specific absences from nine county school districts. Negative binomial regression predicted influenza cases using all-cause and illness-specific absence rates, calendar week, average weekly temperature and relative humidity, using four cross-validations.

Results

School districts reported 2,184,220 all-cause absences (2010-2015). Three one-season studies reported 19,577 all-cause and 3,012 ILI-related absences (2007, 2012, 2015). Over seven seasons, 11,946 confirmed influenza

cases were reported. Absences improved seasonal model fits and predictions. Multivariate models using elementary school absences outperformed middle and high school models (relative mean absolute error (rel-MAE)=0.94, 0.98, 0.99). K-5 grade-specific absence models had lowest mean absolute errors (MAE) in cross-validations. ILI-specific absences performed marginally better than all-cause absences in two years, adjusting for other covariates, but markedly worse one year.

Conclusions

Our findings suggest seasonal models including K-5thgrade absences predict all-age confirmed influenza and may serve as a useful surveillance tool.

Keywords: human influenza, surveillance, prediction, school-aged children

Introduction

Influenza surveillance utilizes multiple data sources, including syndromic indicators, laboratory-confirmed cases, and deaths(1). Non-clinical sources have the potential to complement clinical and laboratory data and improve influenza prediction efforts(2). Student absenteeism is a real-time school-based indicator for influenza surveillance tool. It is advantageous for being widely available in real-time, having minimal reporting delays, and being relatively low cost and is a reasonable proxy for influenza infections since school-age children (5-17 year olds) experience higher infections compared to other age-groups(3) and contribute to household and community-level transmission(4).

Previous studies used school-based surveillance (i.e. absence duration(5) or causes(6, 7)) to identify patterns correlated with influenza- or ILI-related cases, primarily at a city-level, but usefulness of school absenteeism as a surveillance indicator in these studies has been mixed. Using ILI-specific absence duration predicted 2005-2008 outbreaks well in Japan with high sensitivity and specificity(5), but a similar approach using city-level all-cause absences from 2005-2009 had low predictive ability when predicting outbreaks in New York City(6). Absence patterns correlated well with sentinel surveillance in Hong Kong showing similar peaks in absenteeism and ILI consultation and influenza detection rates, but ILI-specific absences had low specificity(8). The varied conclusions of these studies could be from differing school- and absence-types captured, and short surveillance periods, but other types of absence data could have utility.

Grade-specific differences in absences have not been explored as a predictor of influenza but may correlate better to high-risk infections-groups. Given the variation of infection burden and proportion of illness-related absences by age, particular individual school levels and grades may serve as a proxy for these high-risk infection groups. School absenteeism may also be useful for detecting underlying viral changes in transmission. Unusual patterns of school absences arising across different periods of time have also been correlated to detecting changes in influenza A and B viruses(9) and have been attributed to detecting the re-emergence of an influenza B/Victoria antigenic group antigenic group(10). The varied study findings of school absenteeism suggest further assessment is needed.

Here, we evaluated how school absences models predicted weekly confirmed influenza cases in Allegheny County, Pennsylvania over multiple influenza seasons. We compared predictions from all-cause absence models for the 2010-2015 influenza seasons at varying administrative levels. We also compared predictions for individual influenza seasons (2007-2008, 2012-2013, and 2015-2016) from models including all-cause and ILI-related absences from three school-based cohort studies.

Methods

Ethics

Our analyses used only de-identified data. We obtained Institutional Research Board approval from University of Pittsburgh (PRO13100580), Johns Hopkins Bloomberg School of Public Health (IRB #5474), Centers for Disease Control and Prevention (IRB#00000319), and the Allegheny County Department of Health.

Data

Allegheny County Health Department (ACHD) provided virologically-confirmed influenza case data (N=11,946) for 2007-2008, and 2010-2015 influenza seasons. A reported confirmed influenza case was a positive laboratory-confirmed test (i.e., rapid diagnostic test, reverse-transcriptase polymerase chain reaction, or viral culture) reported by hospital emergency departments or sentinel medical providers in Allegheny Country from any individual experiencing influenza-like illness. Weekly all-cause absences and school enrollment data for 2010-2015 came from nine Allegheny County school districts. Six districts provided grade-specific absences (Supplemental Text and Supplemental Table 1). Grade-level served as an age-proxy, since student demographics (i.e., age, gender, or vaccination status) were unavailable. Additional all-cause and cause-specific absences came from three school-based cohort studies (Pittsburgh Influenza Prevention Project (PIPP) during the 2007-2008 season (10 K-5 schools), Social Mixing and Respiratory Transmission in School study (SMART) during the 2012-2013 season (eight K-12 schools), and Surveillance Monitoring of Absences and Respiratory Transmission (SMART^2) study during the 2015-2016 season (nine K-12 schools)). Cohort studies used similar absence collection protocols(11).

Greater Pittsburgh area daily minimum and maximum temperature and relative humidity data came from the National Oceanographic and Atmospheric Administration's National Climatic Data Center(12). We used temperature and relative humidity (a proxy for absolute humidity) given their effects on influenza transmission (i.e., viral dispersal and survival)(13, 14). Allegheny County population data came from US Census Bureau's yearly estimates for 2007, 2008, and 2010-2015(15).

Our primary outcome was weekly confirmed influenza infections reported in Allegheny County during 2007-2008 and 2010-2015 seasons. Influenza infection was defined as any virologically-confirmed case reported by a health provider in Allegheny County during CDC-defined influenza seasons (i.e., 40$^{\text{th}}$ calendar week to the 20$^{\text{th}}$ calendar week of the subsequent year)(16) Weekly influenza cases were the total cases reported each week, excluding cases occurring during school closures (e.g., spring break, federal holidays, weekends).

All-cause absences were defined as a full or partial school day missed for any reason. Cause-specific absences were a full or partial school day missed due to influenza-like illness (i.e., fever (>37C) and either cough, sore throat, runny nose, or congestion). We restricted school absences to periods overlapping the influenza seasons to examine absence patterns during influenza circulation, and excluded weekends, observed federal holidays, and school breaks. Weekly school absences were the total absences reported in one school week (i.e., if no observed holidays, five days in a school week). Weekly absence rates were total absences in a week, divided by the total students enrolled times the number of school days in a given week.

Daily average minimum and maximum temperatures and relative humidity were used to estimate weekly average temperature and relative humidity for each influenza season.

Statistical analysis

We predicted weekly influenza cases over seven influenza seasons using negative binomial regression models. Continuous predictor variables were weekly absence rates (lagged by one-week), calendar week, average weekly temperature, and relative humidity. Models used predictors individually and in combination. The offset term represented the estimated annual Allegheny Country population for 2007 and 2010-2015 influenza seasons. Seasonal variables (calendar week, temperature, and relative humidity) accounted for temporal and climatic variation of influenza. We modeled calendar week, average weekly temperature, and average weekly relative humidity as nonlinear terms using thin-plate penalized splines in generalized additive models (mgcv R package)(17). Models including school- (i.e., elementary, middle and high school) and grade-specific absences (alone and in combination) were evaluated to determine if finer administrative-level absences improved model fits and predictions. From three school-based cohort studies, we compared all-cause and cause-specific absence model performance for single seasons (2007, 2012, and 2015), and pooled over these seasons.

Sensitivity analyses examined absence duration, and lagged influenza, and kindergarten-specific absences. We used one-day and two-day or longer absences to assess the impact of absence duration on weekly influenza predictions from 2010 to 2015. Models used one-day absences, and absences two days or longer individually, together, and in models containing average temperature, relative humidity, and calendar week. We also

assessed weekly influenza predictions from models including one-week lagged influenza cases, and county-level and kindergarten-specific all-cause absences.

We compared nested and non-nested models using Akaike's Information Criterion corrected for small sample sizes (AICc). Decreased AICc signified improved model fits. Two-sided 5% alpha-level determined statistical significance. Analyses used R version 3.1.3 (R Foundation for Statistical Computing, Vienna, Austria, 2016).

Model validation and predictions

Using in-sample data for model training and out-of-sample data for model testing, we used the following four validations: 1) randomly sampled 80% of weeks without replacement; 2) leave out 52 non-contiguous randomly sampled weeks; 3) leave out 20% of randomly sampled schools, and 4) leave one influenza season out (i.e., model training used all but one season and the out-of-sample season was used for model testing) to account for influenzas' seasonal variation. Estimated R^2 used linear regressions of out-of-sample observed influenza cases (outcome) and predicted cases (independent variable). Prediction metrics used mean absolute error (MAE) and relative mean absolute error (relMAE). Mean absolute error was defined as the mean of the absolute value of model prediction errors(18). Relative MAE is the ratio comparing a model's MAE to a reference MAE (i.e., from a model including calendar week, and average weekly temperature, and relative humidity), where relMAE of 1.0 indicated the same prediction error for two models. We visually compared observed and predicted cumulative distributions and time-series of influenza cases.

Results

Characteristics of influenza and school absences

Over seven influenza seasons, 11,946 confirmed influenza cases were reported to ACHD (Supplemental Table 1). Influenza type A predominated most seasons, similar to national patterns(19-24). Overall, 9,350 type A (1,397 A/H3N2 and 1,115 A/H1N1 subtypes) cases, 2,453 type B cases and 143 un-typed cases were reported. The 2011-2012 and 2014-2015 seasons were the lowest (301 cases) and highest (3,150 cases) transmission seasons in Allegheny Country, like national trends. Within seasons, cases peaked in the winter whereas county-level absences varied throughout the year (Figure 1).

During the 2010-2015 seasons, county school districts reported 2,184,200 total absences (Figure1), averaging 6.5 weekly absences/100 students (interquartile range [IQR: 5.6, 7.7]) (Supplemental Table 1). High schools had the highest average absence rates (9.4 weekly absences/100 students, IQR: 8.1,10.8), followed by middle schools (6.3 weekly absences/100 students, IQR: 5.4, 7.7) and elementary schools (5.3 weekly absences/100 students, IQR: 4.0, 6.6). Study schools reported 20,128 all-cause and 3,012 ILI-specific absences among 11,660 students (Supplemental Table 1). The $SMART^2$ study had the highest average weekly all-cause absence rates (2.2 weekly absences/100 students (IQR: 1.8, 2.5)), while the SMART study had the highest ILI-specific absence rates (1.1 weekly absences/100 students (IQR: 0.7, 1.4)).

Influenza predictions using county-level absences

We evaluated negative binomial models of seasonal variables (i.e., calendar week, average weekly temperature, and relative humidity) alone, and including weekly all-cause county-level school absences at one-, two-, and three-week lags. One- and third-week lagged absences had similar model performance (Supplemental Table 2), therefore, we used one-week lagged absences in all models to better reflect influenza's infectious period (i.e. one-week spread)(25). Compared to seasonal models, AICs of in-sample models including calendar week, average weekly temperature, average weekly relative humidity, and one-week lagged weekly county-level all-cause absences either stayed the same or slightly worsened ([**?**] AICc=2, 1, and 0, Table 1), whereas models of calendar week, average weekly temperature, and one-week lagged weekly absences had slightly improved fits ([**?**] AICc=-4, -4, -4, Table 1). For prediction performance, MAEs either stayed the same or decreased when including one-week lagged weekly absences in models of calendar week, average weekly temperature and relative humidity relative to seasonal-only models (relMAE=0.95, 1.0, & 0.95, Table 1).

For individual influenza seasons, weekly-lagged country-level absence multivariate models predicted atypical

seasons poorly, but predicted more typical seasons (i.e., 2013-2014, 2014-2015) with relatively high accuracy (R^2 of 0.91 and 0.57) (Figure 2A). Predicted seasonal peaks were earlier and over-predicted during low transmission seasons (i.e., 2010-2011 and 2011-2012), whereas during high transmission seasons (2014-2015) had later predicted peaks, but of equal magnitude (Figure 2A & 2C). Compared to seasonal models, predicted cases from all-cause absence models varied (either increased or decreased) over the five seasons (Figure 2B), with seasonal peak timing varying most (Figure 2B). Calendar week, average weekly temperature, and absence models varied the most across seasons (Figure 2B). The model containing all seasonal variables and weekly absences had the smallest changes in predicted cases. Lowest MAE models depended on the withheld validation season (Supplemental Table 5). Given the consistently low MAEs of the model including calendar week, average weekly temperature, average weekly relative humidity and school absence, we present results from this model.

Influenza predictions using school-type and grade-specific absences

We compared the performance of different school-types (elementary, middle, and high school) and grade-specific absences in seasonal models. Elementary school models had lower relMAEs compared to middle and high school models across validations (Supplemental Table 6). Given varied model performance across school types, we also considered one-week lagged grade-specific all-cause absences in seasonal models to assess heterogeneity in predictions by grades.

Univariate analyses found K, 1, 2, 3,4th and 5th grade absence models had lower MAEs than (individual) middle school and high school grade-specific absence models, particularly in leave 20% of schools' out validation (Figure 3A). Multivariate grade-specific absence models also had lower MAEs relative to seasonal models across three cross-validations (Figure 3B). We observed consistently lower relMAEs for kindergarten-specific absences (relMAE: 0.91, 0.98. 0.92 in three validations). Overall, middle and high school grade-specific absence models did not decrease MAEs relative to seasonal models, although 8, 9, & 10th grade models in leave 20% of weeks out and 6th grade models in leave 20% schools out had lower MAEs.

We investigated whether absenteeism can be used to create more accurate predictions of virologically confirmed influenza only in school aged children, we built models of virologically confirmed influenza only in those 5-17 years old, rather than of all ages. We found modest improvements in two of three validations when including absences compared to not using absences in models that incorporated week of year, relative humidity and temperature. Predictions were more accurate when predicting virologically confirmed influenza in children then when predicting all ages.

Influenza predictions comparing all-cause and influenza-like illness-specific absences from cohort data

Using school-based cohort studies, we compared the performance of all-cause absences to ILI-specific absences, a better proxy for influenza infection. Because the cohorts had short time-series (i.e., one influenza season), we were unable to examine models containing all seasonal variables and to include average temperature in some models. Multivariate ILI absence models had higher R^2estimates and lower relMAEs than all-cause absence models in analyses using PIPP, 2012-2013 SMART, and pooled absence data (Table 2). From the 2015-2016 SMART^2 data, the all-cause absence model had a lower relMAE (relMAE: 0.59) than the ILI-specific model (relMAE: 2.17), but similar R^2 estimates (Table 2). Pooling across studies, the ILI-specific absence model had a lower relMAE (relMAE:0.99) than the all-cause absence model (relMAE: 1.02), and similar R^2 estimates (all-cause R^2: 0.30 and ILI-specific R^2: 0.37) (Table 2).

Sensitivity analyses

In sensitivity analyses, we found using absence duration did not improve model predictions. One-day absence models and those including both one-day and absences two-days or more had lower relMAEs compared to models containing absences two-days or more (Supplemental Table 9), but predictions from the three model did not substantially vary (Supplemental Figure 2). Evaluation of models including one-week lagged influenza cases found little improvement of model prediction and performance when compared to seasonal models. Higher MAEs were observed for one-week lagged influenza models, and one-week lagged influenza

and absence models but had similar R^2 estimates (Supplemental Table 8). One exception was the one-week lagged influenza model from the leave 20% schools' out validation, which had a lower relMAE (relMAE: 0.97) (Supplemental Table 8). One-week lagged influenza and one-week lagged influenza and kindergarten absence models performed similarly to one-week lagged influenza models in three cross-validations, except in the leave 52-weeks out validation (relMAE: 0.97).

Discussion

We found including school absences in seasonal models improved community-level confirmed influenza predictions over multiple seasons within Allegheny County. All-school absence models subtly improved predictions, reducing MAE by 5% across multiple validations, but school- and grade-specific absence models had better predictions, reflecting underlying age-specific differences in infections. Elementary school absence (K to 5^{th} grades) models decreased MAEs by 1-16% compared to 6-12^{th} grades, suggesting younger student absences were illness-related and older children's absences were non-influenza and non-illness related. From school cohort data, ILI- and all-cause absences performed better in single season (2007-2008 and 2012-2013) validations and when pooled across seasons. Elementary school, K-5^{th} grade-specific all-cause absences, and potentially ILI-specific absences, may serve surveillance indicators for the larger community.

Compared to seasonal models, those including all-cause absences improved MAE and R^2 estimates, and suggests that after accounting for seasonal factors, school absences improved influenza predictions. Our analysis is one of few using weekly all-cause absences at various administrative levels (i.e., school type and grades) to predict influenza. Whereas other studies used cause-specific absences to detect elementary school influenza outbreaks(6), ours evaluated how different school and grade all-cause absences performed as predictors. As evidenced by higher R^2 and lower relMAEs from elementary school absence models, absences from younger school-aged children better reflect infections during the influenza season and are a proxy to the younger age groups that experience higher infections and increased susceptibility(5, 25, 26). In contrast, middle and high schools' absences were noisier prediction signals, possible because older students had more non-influenza related absences (consistent with the overall higher absenteeism rates observed in these schools over time). Lower relMAEs from lower individual grade (K-5^{th} grades) absence models from multiple validations further support our findings. Hence, elementary school absences could be useful for influenza surveillance.

ILI-specific absences predicted influenza better than all-cause absences when evaluating predictions from weekly all-cause and ILI-specific absence models (using school-based cohort studies), based on lower MAEs and higher R^2 for specific seasons and when pooled. Other studies also found ILI-specific absences were a proxy for influenza when evaluating vaccine impacts(27), suggesting ILI-specific absences likely capture actual influenza infections. We could not conduct cause-specific absence surveillance for more than one influenza season for each study nor could we perform school-type and grade-specific comparisons of all-cause and ILI-specific absences due to small time-period, but these may also be important predictors of influenza incidence.

Our study has some limitations. We did not evaluate our predictions during the 2009 pandemic because our county absence data were either limited to single seasons, or available after 2009 because participating schools' electronic absence surveillance began after 2009. Similarly, cohort studies were funded for and conducted during the 2007, 2012, and 2015 seasons, therefore we could not assess predictions during the 2009 pandemic. In the school-based cohort studies, not all absences were identified due to challenges contacting parents regarding absences and our studies may underestimate the number of all-cause absences, and possibly, ILI-specific absences. Our predictions used school-based data from school districts within Allegheny County only, therefore our results may not be generalizable to influenza transmission in other US counties. Additional data from other Pennsylvania counties or a representative sampling from other state counties would improve the generalizability of our predictions.

Recently, others, like those participating in the CDC FluSight Challenge – an influenza prediction competition – have used climate data, past influenza incidence and other data streams in recent efforts. In the CDC

FluSight Challenge, external research teams predict weekly influenza cases, and evaluation metrics include the mean absolute scaled error, a measure of forecast accuracy(28, 29). Our MAE decreased by 5% when using county-level all-cause absences models and is equivalent an additional 8 weeks of data included in a nowcast model, like those used in the FluSight Challenge. This equates to a 5% reduction in mean absolute scaled error(30). Our results suggest that models including lower grades' absences may improve predictions, as seen by the 10% MAE decrease, and may improve predictions more when incorporated into ensemble models, like those used in FluSight(29).

Our findings suggest models using absences of younger students improves predictive performance. Real-time, day-to-day absence data are easy to collect, readily available in many schools, and can provide more accurate predictions than other surveillance mechanisms reliant on virologic confirmation, and susceptible to laboratory testing delays. Future studies could apply absence data to other prediction methodologies, like ensemble methods and machine-learning algorithms, which may improve prediction accuracy and identify absence-related patterns not considered here. We demonstrate grade-specific all-cause absences predict community level influenza one-week forward, when influenza- or cause-specific absences are unavailable and suggest elementary school or lower grade absenteeism during the influenza season can reflect influenza circulation. Using school indicators can inform influenza surveillance and control efforts, including annual vaccination; antiviral treatment or prophylaxis; and promotion of everyday preventive measures (i.e., staying home when sick, respiratory hygiene, and hand hygiene) to reduce school- and community-level influenza transmission.

References

1. CDC. Overview of Influenza Surveillance in the United States Atlanta, GA: CDC; 2016 [cited 2016 May 26]. Available from: http://www.cdc.gov/flu/weekly/overview.htm.

2. Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. J Infect Dis. 2016;214(suppl_4):S380-s5.

3. Worby CJ, Chaves SS, Wallinga J, Lipsitch M, Finelli L, Goldstein E. On the relative role of different age groups in influenza epidemics. Epidemics. 2015;13:10-6.

4. Monto AS, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. Epidemiol Infect. 1993;110(1):145-60.

5. Sasaki A, Hoen AG, Ozonoff A, Suzuki H, Tanabe N, Seki N, et al. Evidence-based tool for triggering school closures during influenza outbreaks, Japan. Emerg Infect Dis. 2009;15(11):1841-3.

6. Egger JR, Hoen AG, Brownstein JS, Buckeridge DL, Olson DR, Konty KJ. Usefulness of school absenteeism data for predicting influenza outbreaks, United States. Emerg Infect Dis. 2012;18(8):1375-7.

7. Tan L, Cheng L, Yan W, Zhang J, Xu B, Diwan VK, et al. Using daily syndrome-specific absence data for early detection of school outbreaks: a pilot study in rural China. Public Health. 2014;128(9):792-8.

8. Cheng CK, Cowling BJ, Lau EH, Ho LM, Leung GM, Ip DK. Electronic school absenteeism monitoring and influenza surveillance, Hong Kong. Emerg Infect Dis. 2012;18(5):885-7.

9. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection, New York City. BMC Public Health. 2005;5:105.

10. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol. 2013;9(10):e1003256.

11. Stebbins S, Cummings DA, Stark JH, Vukotich C, Mitruka K, Thompson W, et al. Reduction in the incidence of influenza A but not influenza B associated with use of hand sanitizer and cough hygiene in schools: a randomized controlled trial. Pediatr Infect Dis J. 2011;30(11):921-6.

12. (NCEI) NCfEI. Climate Data Online 2016 [cited 2016 May 25]. Available from: http://www.ncdc.noaa.gov/.

13. Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M. Absolute humidity and the seasonal onset of influenza in the continental United States. PLoS Biol. 2010;8(2):e1000316.

14. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathog. 2007;3(10):1470-6.

15. Census U. Quickfacts: Allegheny Country, Pennsylvania 2016 [Available from: http://www.census.gov/quickfacts/table/PST045215/42003.

16. CDC. Influenza (Flu): CDC; 2013 [cited 2016 May 26]. Available from: http://www.cdc.gov/flu/about/season/flu-season.htm.

17. Wood S. Generalized Additive Models: An Introduction with R (2nd ed.). : CRC Press; 2006.

18. Reich NG, Lessler J, Sakrejda K, Lauer SA, Iamsirithaworn S, Cummings DA. Case study in evaluating time series prediction models using the relative mean absolute error. Am Stat. 2016;70(3):285-92.

19. Update: Influenza activity–United States and worldwide, 2006-07 season, and composition of the 2007-08 influenza vaccine. MMWR Morb Mortal Wkly Rep. 2007;56(31):789-94.

20. Update: influenza activity - United States, 2011-12 season and composition of the 2012-13 influenza vaccine. MMWR Morb Mortal Wkly Rep. 2012;61(22):414-20.

21. Influenza activity–United States, 2012-13 season and composition of the 2013-14 influenza vaccine. MMWR Morb Mortal Wkly Rep. 2013;62(23):473-9.

22. Appiah GD, Blanton L, D'Mello T, Kniss K, Smith S, Mustaquim D, et al. Influenza activity - United States, 2014-15 season and composition of the 2015-16 influenza vaccine. MMWR Morb Mortal Wkly Rep. 2015;64(21):583-90.

23. Davlin SL, Blanton L, Kniss K, Mustaquim D, Smith S, Kramer N, et al. Influenza Activity - United States, 2015-16 Season and Composition of the 2016-17 Influenza Vaccine. MMWR Morb Mortal Wkly Rep. 2016;65(22):567-75.

24. Epperson S, Blanton L, Kniss K, Mustaquim D, Steffens C, Wallis T, et al. Influenza activity - United States, 2013-14 season and composition of the 2014-15 influenza vaccines. MMWR Morb Mortal Wkly Rep. 2014;63(22):483-90.

25. Medina RA, García-Sastre A. Influenza A viruses: new research developments. Nat Rev Microbiol. 2011;9(8):590-603.

26. Hayward AC, Fragaszy EB, Bermingham A, Wang L, Copas A, Edmunds WJ, et al. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. Lancet Respir Med. 2014;2(6):445-54.

27. Kjos SA, Irving SA, Meece JK, Belongia EA. Elementary school-based influenza vaccination: evaluating impact on respiratory illness absenteeism and laboratory-confirmed influenza. PLoS One. 2013;8(8):e72243.

28. CDC. Epidemic Prediction Initiative 2018 [cited 2018 Jan 24]. Available from: https://predict.phiresearchlab.org/post/595d3c4545e6b6190e8f183c.

29. Biggerstaff M, Alper D, Dredze M, Fox S, Fung IC, Hickmann KS, et al. Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge. BMC Infect Dis. 2016;16:357.

30. Brooks LC, Farrow DC, Hyun S, Tibshirani RJ, Rosenfeld R. Flexible Modeling of Epidemics with an Empirical Bayes Framework. PLoS Comput Biol. 2015;11(8):e1004382.

Tables

Table 1. Fit and Performance of Negative Binomial Models of Seasonal Variables Including and Excluding One-week Lagged County-level All-cause School Absence Rates to Predict Weekly Confirmed Influenza Cases in Allegheny County, Pennsylvania During the 2010-2015 Seasons

Model Validation	Model Validation	Leave 20% randomly sampled out data (n=1
Model^{+}	**Variables**	**D.f.**
1 (Ref.)	Week, temperature, RH	8.5
2	Week, temperature, all-cause absence rates	7.3
3	Week, RH, all-cause absence rates	7.9
4	Week, temperature, RH, all-cause absence rates	8.8

$^{+}$Each model used negative binomial regression and used generalized additive models to estimate degrees of freedom for non-linear (i.e. spline) variables. $^{++}$Changes in AICc and relMAE compared all models to the reference (model 1), a seasonal variables-only model that contains calendar week, average weekly temperature, and average weekly relative humidity. Abbreviations: [?]AICc, change in Akaike's Information Criterion corrected for small sample size; RH, relative humidity.

Table 2. All-cause and Cause-specific Absences Model Performance Using Three School-based Cohorts' Data to Predict Confirmed Influenza Cases in Allegheny County, Pennsylvania, USA During the 2007-2008, 2012-2013, and 2015-2016 Influenza Seasons

Flu Season	Cohort	Model$^{+,\ ++}$	In-sample AICc ([?])	R^2	MAE	Relative M
2007-2008	PIPP	Week-only	183 (0)	0.97	8.9	1.0 (Ref.)
		All-cause absence	215 (32)	0.44	158.6	17.8
		ILI-specific absence	185 (2)	0.49	45.1	5.07
2012-2013	SMART	Week-only	161 (0)	0.93	14.4	1.0 (Ref.)
		All-cause absence	187 (26)	0.98	11.2	0.78
		ILI-specific absence	174 (13)	0.99	8.6	0.60
2015-2016	SMART^2	Week-only	204 (0)	1.0	8.6	1.0 (Ref.)
		All-cause absence	214 (10)	0.82	5.0	0.59
		ILI-specific absence	206 (2)	0.84	18.6	2.17
Pooled analysis	PIPP, SMART, SMART^2	Week-only	664 (0)	0.35	51.6	1.0 (Ref.)
		All-cause absence	665 (1)	0.30	52.4	1.02
		ILI-specific absence	667 (3)	0.37	51.5	0.99

$^{+}$The week-only model included only week of the year and absence models included weekly lagged absence rates from the previous week, week of the year, and average temperature.$^{++}$SMART models included weekly lagged absence rates and week of the year. ^{c}Cross-validation used leave 20% of schools out. Abbreviations: [?]AICc: change in corrected Akaike's Information Criterion; ILI: influenza-like-illness; MAE: mean absolute error; PIPP: Pittsburgh Influenza Prevention Project; SMART: Social Mixing and Respiratory Transmission in Schools study; SMART2: Surveillance, Monitoring of Absences & Respiratory Transmission Study; R^2: coefficient of determination.

Figure legends

Figure 1. Weekly reported virologically-confirmed influenza cases, and all-cause and influenza-like-illness (ILI) specific absences in Allegheny County, Pennsylvania, USA, during influenza seasons from 2007 to 2015 . Surveillance of influenza cases during each influenza season in Allegheny

County occurred from the 40th week of one year to the 20th week of the following year (solid black lines). All-cause absences were collected for the entire school year for each school district, and data were restricted to their respective influenza seasons. Nine school districts within Allegheny contributed to weekly counts of all-cause absences. Additionally, all-cause and ILI-specific absences were collected during independent influenza seasons for three separate studies: 2007-2008 (PIPP study), 2012-2013 (SMART study), and 2015-2016 (SMART^2 study). Absence surveillance data were not collected during the 2008-2009 or 2009-2010 influenza seasons. White space on the x-axis reflects periods when data were not collected for this analysis, whereas black lines on the x-axis (in negative y values) indicate time periods when data is available. The Allegheny County student population averaged 43,636 students across the nine school districts, comprising 122 schools (57 elementary, 20 middle, and 18 high schools, and 24 charter/independent schools). County-level data for 2010-2011 season were not available for 3 school districts.

Figure 2. Four model predictions of confirmed influenza in Allegheny County using leave one season out validations for the 2010 to 2014 influenza seasons. Model predictions of four negative binomial models (calendar week, average weekly temperature, average weekly relative humidity (red), one-week lagged county-level all-cause absences, temperature, and week model (yellow), one-week lagged county-level all-cause absences, relative humidity, and week model (blue), and one-week lagged county-level all-cause absences, temperature, relative humidity, and week model (purple)) using leave-one-season-out validation approaches showing model predictions compared to observed virologically-confirmed influenza cases (black line) in Allegheny County, Pennsylvania, USA, A) weekly counts during each of the 2010-2011 to 2014-2015 influenza seasons, B) the change in predicted cases using modeling including absences compared to a seasonal model excluding absences (red), and C) the cumulative proportions of predicted and observed influenza cases for each season. R^2 was obtained using a linear regression, where the observed cases from the left-out season are the dependent variable and the independent variable were predicted cases from a negative binomial model of week-lagged county-level all-cause absences, relative humidity, temperature, and calendar week.

Figure 3. Mean and relative absolute errors for predictions using grade-specific absence models to predict influenza in Allegheny County Pennsylvania from 2010 to 2014 influenza seasons. Mean absolute errors were estimated from univariate grade-specific weekly absences models (A), and the relative mean absolute error compared models of grade-specific weekly absences, week of the year, average weekly relative humidity, average weekly temperature to models of calendar week, average weekly relative humidity, and average weekly temperature (B). Colors reflect the three different school types: red is elementary school, green is middle school, and blue is high school. Solid black line refers to a relMAE of 1, where mean absolute errors of the grade-specific absence models and models excluding absences are the same.