PI: Derek Cummings
Study Title: Monitoring cause specific absences to estimate influenza transmission (SMART2)
IRB No.: IRB00005474
Date: 8/27/2014

1. Aims/Objectives/Research Question/Hypotheses

This project will lead to improved understanding of how diseases such as influenza spread among children in school and between children and the community, will lead to more accurate forecasting of influenza incidence during seasonal influenza epidemics and improved understanding of the correlation of influenza incidence and the incidence of other respiratory transmitted pathogens. It will also inform the creation of transmission models of respiratory disease and help us understand the impact of respiratory diseases on schools on absenteeism.

The primary goal of SMART2 is to improve prediction and tracking of influenza activity in entire communities. The research question investigates the connection between influenza among school children and outbreaks, epidemics, and pandemics in the wider community. Our hypothesis is that we can create a predictive model of community influenza by monitoring absenteeism in schools. This model, if successful, could afford advance warning of disease outbreaks and inform prophylactic strategies.

Specific Aim 1: Monitor absenteeism (SA1)
We will monitor student absenteeism in target K-12 schools using electronic reporting systems.

Specific Aim 2: Develop a surveillance system to detect and determine the symptoms of illness associated with absences of school children from 3 school districts in and near Pittsburgh, PA (SA1)

> Specific Aim 2A: Obtain absences from established school systems and develop multiple systems to contact parents to elicit a more specific cause of the absence and symptoms where the absence was through illness. In two previous projects in which we contacted parents to determine the cause of absence, we have found that contact by only telephone is limited in reaching many parents. We will attempt to contact parents and guardians of school children through telephone, email and text methods. These systems will be integrated within the existing absence follow-up of the school systems.
>
> Specific Aim 2B: Enroll a subcohort of 360 children and their families to be contacted weekly during an 12 week period to determine if any respiratory illness has occurred in the past week. Families will be recruited into the cohort study from 6 schools participating in study that are providing individual level data.

Specific Aim 3: Determine etiology of illness and detect respiratory illness both

associated and non associated with school absence (SA3)

Specific Aim 3A: Obtain nasal swabs from school children who are leaving school due to illness consistent with influenza or another respiratory pathogen was indicated by a parent or guardian and use multiplex RT-PCR to determine the presence/absence of influenza and a panel of additional respiratory viruses.

Specific Aim 3B: Obtain nasal swabs from school children who are returning to school due after an absence where illness consistent with influenza or another respiratory pathogen was indicated by a parent or guardian, and use multiplex RT-PCR to determine the presence/absence of influenza and a panel of additional respiratory viruses.

Specific Aim 3C: Conduct surveillance for respiratory illness in a sub cohort of schoolchildren and use multiplex RT-PCR to determine the presence/absence of influenza and a panel of additional respiratory viruses from nasal swabs collected when children return to school after an illness detected through follow up of the sub-cohort.

Specific Aim 4: Develop predictive model to forecast temporal patterns of incidence and strain distributions of both laboratory confirmed influenza infections (medically attended or hospitalized) and cases of influenza like illness within surrounding communities using cause-specific absences available from publicly available data sources or sources available to us through agreement to obtain non-personally identifiable data and infection data from our school cohorts.

Specific Aim 4A: Obtain and process surveillance data on laboratory confirmed influenza outcomes from the University of Pittsburgh Medical Center (publicly available surveillance data), Allegheny County Health Department (publicly available surveillance data), Pennsylvania Department of Health (publicly available surveillance data) and US Centers for Disease Control (publicly available surveillance data).

Specific Aim 4B: Determine the performance of multiple statistical models in predicting outcome data from the broader community based passive surveillance data sets, using a likelihood based methodology

Specific Aim 4C: Compare prediction algorithms based on cause specific absence data to alternatives, including algorithms based on historical passive surveillance data and climate data.

2. Background and Rationale

School-age children are thought to play a central role in spreading pandemic viruses. Children experience disproportionately higher rates of infection, shed influenza virus for approximately twice as long as adults, and have important social mixing patterns that are thought to lead to higher rates of contact than the rest of the population. Incidence during pandemics has been seen first among school-aged children and was shown to be important in maintaining tramission during the spring wave of the 2009 H1N1pdm virus.

The central role that children play in the transmission of influenza has lead to the suggestion that surveillance should be more focused on school-age children. The availability of absentee reports from an ever-increasing number of schools utilizing electronic reporting systems increases the potential of using absentee data to track and predict the temporal progression of influenza incidence in communities.

However, there are limited and conflicting studies that have assessed the utility of cause-specific school absence data to predict community wide influenza activity. Studies pairing passive surveillance data, absentee surveillance and laboratory confirmation of illness associated with school absence have rarely been performed. The impact of incidence due to other respiratory pathogens on the predictive capacity of cause-specific school absence has not been studied. Understanding the relationship and possible interactions between other respiratory infection incidence, school absence and community influenza activity could substantially improve predictive models. This, in turn, would improve the guidance of prophylactic measures available to schools during disease outbreaks.

We propose to use traditional time series modeling approaches and Hidden Markov Models to model the temporal dynamics of influenza activity in the community as a function of cause-specific school absence and other covariates. We will also use these models to predict the future course of influenza in the school-aged population as a function of cause specific school absences in earlier time periods. We also propose to use methods to detect spatial-temporal signatures of incidence that may be associated with the onset of the influenza season in communities.

The primary goal of the proposed effort is to improve prediction and tracking of influenza activity in entire communities. We aim to do this by providing information on the conduct of optimal surveillance in school-based populations to improve prediction of influenza activity in communities and development of novel statistical methods that can be extended to populations beyond those studied here. Early warning of the timing and extent of an influenza epidemic can substantially improve the response and preparation of healthcare services, both in gearing up to accommodate predicted patient surge, and in postponing or rescheduling routine non-essential services or consultations. Community education can also be improved with early knowledge of influenza season timing, including education about early, appropriate vaccination.

Our group has already begun work on trying to address the aforementioned gap in knowledge of social mixing patterns of school age children. In a previous projects, co-Investigators of this proposal have conducted surveillance for infleunza like illness and investigated social contact patterns and surveillance in two school districts in Pittsburgh the last 5 years. The proposed study in this application will be an extension of our previous work.

3. Participants

The primary population of interest to our study is the students of three school districts in Allegheny County: (1) Fox Chapel School District (3000 students in 7 schools covering K-12) (2) Propel Charter School District (3000 students in 9 schools covering K-12) and (3) Canon-

MacMillan School District (5000 in 11 schools covering K-12). The leaders of these three school districts have agreed to collaborate with us on the proposed research.

In order to be included in the surveillance part of the study described in the last paragraph, potential participants must be enrolled in one of the chosen schools at Fox-Chapel School District, Cannon-McMillan School District, or Propel School District. To participate, parents must decide not to opt out and children of legal age must decide to not opt out. The opt out process will be conducted each year of the study, providing all participants the opportunity to opt out of the study in each year. Those opting out once can contact the study to opt in, but will not be contacted again in each year if opting out in any previous year.

We expect to obtain absence data on all students in the three school districts covering 11,000 students. From each of these schools, we will obtain the number of absences by school, grade, class and day. We will not be collecting personally identifying data from these schools but only aggregate counts. We will select 6 schools (covering K-12 grades, 2 from each school district) for enhanced surveillance that will require obtaining individual absence reports and personally identifying information. The school districts and school leadership will determine which schools are interested, willing and able to participate. We will follow-up with the child's parents to determine the etiology of illnesses associated with absences.

We will enroll an additional cohort of up to 360 children and their household members into an active surveillance cohort (total sample size ~1440). Members of these households will be contacted weekly during the respiratory disease surveillance season to detect respiratory illness and swabbed upon respiratory illness even if not associated with school absence. Households will be followed for 12 weeks. A parent, guardian or other member of the household over 18 will be asked to provide information on illnesses in the family each week. All members of the household may be asked to provide a nasal swab when another member of the household reports illness. The households of all participating schools will be asked if they would like to participate as cohort households. Parents and guardians in households that have not opted out will be sent a flyer ("Cohort flyer – version 9.docx" included in original submission) to determine if they would like to participate. Cohort households will be selected on a first come, first served basis from these volunteers. If they would like to participate, we ask those parents to contact us to participate. Households participating in one year can also participate in the next year of surveillance but enrollment will be done independently in each year.

To participate in this aspect of the study (the household cohorts), parents and household members over 18 must consent to participate and children must provide assent. Pennsylvania law states that foster children should not be included in research unless it is of direct benefit to them. So we will ask schools to exclude these children.

4. Study Procedures

Study Duration
The study will be implemented for three years, during the flu season from October to May. Each year, publicly available surveillance data will be used to determine the timing of the 12

weeks of intensive surveillance of the sub-cohort to coincide with a period of elevated influenza incidence.

Detection of Absences

Participating schools will provide same day name-based data on absent students during influenza season. In two school districts, lists are electronically available on the same day, while in a third, hard copy of written lists are prepared each day. This list will be used to identify families to be contacted in order to determine the reason for absence. All absences will be entered into a database along with the status of follow-up including failure to follow-up or failure to contact family members.

To be compliant with FERPA regulations, school districts will not provide absence information on students who have been opted out of the study.

We have created database systems for this purpose already for a previous study that are maintained on encrypted drives at computers at the University of Pittsburgh Graduate School of Public Health. Contact will be attempted by multiple means including telephone, email and SMS. Initial contact letters will be sent home to parents to describe the study and to permit opt-out of the study. These initial contact letters will be sent to the parents of newly enrolled or eligible students at the start of each season. Schools have agreed to provide us with telephone numbers, email addresses and mailing addresses for parents of students. Parents will also have the option of filling out a survey online or by phone.

Determining circumstances of absence and symptoms associated with illness

Study staff will contact the homes of individuals who do not opt out of the study. Parents of children in these schools are routinely contacted when children are absent in order to determine the cause of illness for administrative purposes. Students who experience symptoms of influenza like illness will undergo nasal swabs for influenza (RT PCR) at school in a timely and convenient manner or when they return to school after an absence associated with symptoms of influenza like illness. We will have a database that includes a list of students to be followed up when they return to school that the staff will use to identify students that will be asked to provide assent for swabbing. We used this system effectively in a previous study. Students who have influenza like illness will be asked to come to the nurse's office (accompanied by study staff), assent to the study and provide a nasal swab. For those students who were not present at school when ILI symptoms were present, students will be swabbed upon their return from absence associated with ILI symptoms. In our previous study, this process was completed in less than 5 minutes on average. Students will be asked to provide assent. Separate assent scripts have been included for ages 5-6, 7-12 and 13+. Children who report to the school nurse will also be asked about their symptoms by a school nurse using the same illness reporting form and asked to assent to a swab if meeting the study definition of influenza like illness.

School nurses will be given the option of performing swabbing if they prefer, but otherwise

swabbing will be performed by study staff. Both school nurses and study staff will be given training in conducting sample collection.

Creation of sentinel sub-cohort

In order to determine the cause of respiratory illness that are not associated with school absence and assess the efficiency of our surveillance of the larger school population, we will enroll 360 children and their families into a sub-cohort that will act as a sentinel population and will be contacted weekly to determine if any respiratory illness has occurred in the past week and if there has been any school absence associated with illness. A cohort of 360 students and their household members will be enrolled each year. Households participating in one year can participate in another study year, but enrollment will be performed separately in each year.

A parent, guardian or other member of the household 18 or over will be requested to fill out a short web-based survey weekly describing any symptoms experienced in the past week by any household member. We expect this will take less than 5 minutes. Participants will be reminded by telephone if surveys are not completed. If a respiratory illness is identified then students will be asked to report to the school nurses office for a nasal swab to be obtained. Members of the household will be asked to collect a nasal self-swab using a kit that will be sent to participants in the household cohort who experience an influenza like illness. Each of the participants will be asked to provide information on illnesses individually to parent or guardian who will fill out the web-based form. We will encourage family members to fill the survey out together.

Households from all participating schools will be asked to participate as cohort households. Households will be selected on a first come first served basis until 360 households are enrolled.

Survey of students in absence surveillance schools

Students (and their parents) will respond to a survey about their illness associated with an absence by telephone. We expect the survey to take less than 5 minutes.

Survey responses completed by either responding to a telephone survey will be registered in a single database so that students are not contacted multiple times.

Web-based survey of household members in sentinel sub-cohort

We will ask household members in our 360 household sentinel cohort to fill out a web-based survey of their symptoms during the past week for 12 weeks of follow-up per respiratory disease season (to be filled out by parent or guardian or member over 18 in each household). Report of illness meeting our case definition of respiratory disease (Fever over 37.8 C and either cough or sore throat) will be referred to study staff for acquisition of a nasal swab at schools (for students) or at home for household members (by self swab). Families enrolling in the cohort study will be given a $20 gift card for enrolling in this study. Each week when they provide information, they will be entered into a drawing for a $200 gift card. Families

will be given an additional $20 gift card if they complete at least 5 of the first 6 weekly surveys, and another $20 gift card if they complete at least 5 of 6 of the weekly surveys in weeks 7-12 of the study period. All gift cards will be sent at the conclusion of the 12 weeks.

We have included the web-based forms as an appendix ("cohortweeklyillnessreport_SMAR.pdf") as well as the intake questionnaire that will ask participants over the phone about a larger number of covariates than answered each week ("cohortintake_SMART2.pdf"). Parents and guardians in households that have not opted out will be sent a flyer ("Cohort flyer – version 9.docx" included in original submission) to determine if they would like to participate. Those that do want to participate will be asked to call our study staff. An appointment will be made to for our study staff to call and ask the questions on the intake questionnaire ("cohortintake_SMART2.pdf"). We expect that this questionnaire will take 10 minutes.

Measurement of vaccine status and other demographic covariates
Parents of school children in surveillance schools will be surveyed at the end of the school year to determine their vaccination status of their child and other members of the household in the current and past seasons for a subset of individuals in the school. Surveys will be conducted by phone. We will also obtain demographic information about their household and the inhabitants of their household.

Obtaining swabs from surveillance school subset
For a subset of schools for which we will obtain syndrome descriptive information on the cause of absences, we will obtain nasal swabs from children who present to the school nurse with ILI based on a screening questionnaire. Children who return to school after an absence where ILI is indicated will also be tested. Testing will be done after each absence for ILI and specimens will be collected as soon as possible and after up to seven school days of absence. The sensitivity of RT-PCR to detect influenza infections and other respiratory infections will depend upon the time of sample collection in relation to the onset of symptoms. Generally, samples collected shortly after symptom onset have been found to have the highest viral load

Obtaining nasal swabs from sentinel sub-cohort
For the students in our 360 child sentinel cohort, we will obtain nasal swabs for any respiratory illness that meets either of our two case definitions, targeting a higher sensitivity for respiratory infection (two symptoms of either cough, runny nose, sore throat, congestion or fever plus either cough or sore throat) regardless of whether or not the illness is associated with a school absence. Nasal swabs will be obtained at the earliest time possible at the school that the cohort member attends. Nasal swabs will be obtained at the nurse's office for students enrolled in the household cohort. A self swab kit will be delivered to the home by mail for household cohort participants who meet our case definition in this paragraph. Participants are asked to collect a nasal swab and return this using the kit by mail to our study staff.

Swab process
Staff will utilize sterile foam swabs to collect a mucus specimen from the noses of participants; the swab will be placed in sterile transport media and transported via cooler to the University of Pittsburgh Medical Center Clinical Virology Laboratory for testing using

RT-PCR. Tests will be for both influenza A and B, and will include subtyping of H1 and H3 influenza A viruses. RT/PCR is considered the "gold standard" for detection of influenza, with reported selectivity and sensitivity rates in the high 90%. No tests will be repeated, as participants will not be reswabbed.

Determining viral etiologic agent of school absence

Nasal swabs will be placed in viral transport media (Remel Microbiology Products, Lenexa, KS). Nasal swabs will be maintained at 4° C and transported to the laboratory for testing within 72 hours. Samples will be tested using multiplex PCR (Luminex, Austin, TX) for both influenza A and B, and included subtyping of H1 and H3 influenza A viruses. Testing will also be done for respiratory syncytial virus (RSV A,B), parainfluenza viruses 1-4 (PIV 1-4), human metapneumovirus (HMPV), human rhinovirus (HRV), adenovirus (ADNO B,C,E), and coronavirus (CV 229E,NKL63, HKU1, OC43).

Laboratory Testing

The study will monitor influenza activity information from Allegheny County health Department, PA Dept. of Health, UPMC virology lab and other sources.

During influenza season, study staff will monitor absentee data provided by the school. Parents will be contacted to determine if their student has influenza like illness (ILI). Those with ILI (fever of at least 37.8 C and either cough or sore throat) will be tested upon return to school. Students presenting to the school nurse with ILI will also be tested.
Trained project staff will be deployed in the schools to determine children who have been absent from school, ask them to go to the nurses office, where children will be confidentially swabbed, without risk of loss of confidentiality to other students. We will work cooperatively with the school in all instances.

Staff will utilize sterile foam swabs to collect a mucus specimen from the noses of participants; the swab will be placed in sterile transport media and transported via cooler to the University of Pittsburgh Medical Center Clinical Virology Laboratory for testing using RT-PCR. Tests will be for both influenza A and B, and will include subtyping of H1 and H3 influenza A viruses.

Testing will be done at the conclusion of the study year to batch all study results. Results will not be provided back to students and/or their parents.

Sample Size

To estimate the sample size necessary to build effective predictive models, we have fit time-series models to simulated data in order to explore the ability to fit models to data collected on influenza incidence in our school aged cohorts. We did this by assuming a temporal shift between the influenza dynamics in different age groups, in an age-specific model of influenza with coupling between age classes. The temporal shift in this case was created by using different seasonalities of influenza in the two populations. We used information from the SMART study to parameterize a 5-age class model of influenza and initiated influenza in the youngest age class in 3 simulated influenza seasons. We then sampled 500, 1000, 1500,

2000, 2500, and 3000 children from our simulated population (of 100,000) in the lowest two age classes and sampled 5% of cases from the two oldest age classes (with a population of 100,000). We then fit seasonal autoregressive models with 5 time-specific seasonal factors on transmission to the two data streams and compared the performance of two individual models and a shared parameter model. We performed this procedure 1000 times. The two models were determined to be significantly different by this procedure 55%, 68%, 75%, 82%, 88% and 89% of the time given the respective samples of 500, 1000, 1500, 2000, 2500 and 3000. Based upon these results, we chose a targeted sample size of 2000 to 2500, roughly the expected number in our 6 schools. This study will be conducted over multiple seasons and so we expect some amount of turnover in our schools. The total sample size of individuals for which we will obtain individual level data is 4000 school children in our selected schools. We will also ask questions of parents of these individuals and thus expect to contact, conservatively, 8000 parents of these children (assuming two parents per child are contacted over the course of the study)

The sample size of the household cohort (360 households) was determined by calculating the required sample size to estimate the proportion of individuals experiencing influenza virologically confirmed illness within a desired level of precision. We estimate that 15% of individuals in this population will have an influenza virologically confirmed infection during the 12 week period. We would like to estimate this proportion with an absolute range of +/- 2 %. We assume an alpha of 0.05 and power of 0.80. To do this we require 1225 individuals. Assuming that each household has on average 3.6 individuals, we conservatively estimate 360 households.

The household cohorts will come from these populations (for the school children and parents) however, household members who are not parents nor students in our selected schools add additional numbers to the total number of participants for which identifiable demographic data will be collected. This additional number is estimated to be 360-720 additional household participants.

Analysis Plan

We are interested in examining the relationship between cause specific school absenteeism and multiple outcomes derived from surveillance data sets. Outcomes fall into three types: (1) counts (2) weeks at which certain thresholds of the influenza season will be reached and (3) distributions. Each one of these outcomes will require slightly different methods. Below we describe methods used for each including negative binomial regression, regression trees, Hidden Markov models, receiver operating characteristic curves and simple transmission models.

Outcomes: Counts

The primary outcome of this work will be to predict the temporal course of influenza incidence in communities using cause-specific absentee data in schools.
Specific Counts of Interest: (1) Weekly Counts of influenza isolates collected from (a) Pittsburgh (b) Allegheny County (c) Pennsylvania. (2) Weekly counts of influenza isolates collected from each of these locations among those in each 5 year age class of the population (3) Influenza hospitalizations (4) Age specific hospitalizations (5) cause-specific absenteeism

in future time periods.

Each of these counts can be calculated for multiple temporal aggregations or with temporal smoothing.

Outcomes: Thresholds and the timing of peaks

One goal of disease surveillance is the early detections of periods of high incidence in particular disease areas, i.e., outbreak detection. School-based cause-specific absentee surveillance may be useful in predicting when community wide surveillance measures cross particular thresholds. We would be interested in predicting the time of multiple potential thresholds. Times of interest include the week of the influenza season when cumulative influenza isolates measured in the community cross some small threshold, perhaps 5% of the average cumulative number of influenza isolates obtained over the past 5 years. This timepoint could be used to inform when the initiation of the influenza season will occur. We may also be interested in predicting the week when the peak number of influenza isolates in the community will be obtained. Finally, we may be interested in predicting when isolates will cease to be obtained or when a seasonal cumulative total will be exceeded, for example the time when 95% of all influenza isolates for a particular season will be obtained, in order to provide guidance to public health authorities on when control measures may be relaxed.

Outcomes: Distribution of isolates

Among secondary outcomes of interest are the strain distributions of community-based isolates and the age distribution of community-based isolates and hospitalizations.

Outcomes: Spatial-temporal dynamics of influenza infections

Another metric of influenza transmission in a community besides the count of isolates at different time points is the growth in isolates as measured by the effective reproductive number of isolates of influenza. This outcome can be estimated with various degrees of temporal aggregation or smoothing.

Outcomes: Non-influenza respiratory outcomes

Surveillance data is available on several other respiratory pathogens at the community level including RSV and parainfluenza I-III. We will use the methods described below to predict community-wide isolates of each of these pathogens. Similarly, we expect to detect several other respiratory pathogens by using the multiplex RT-PCR.

Statistical Methods

Negative binomial regression

We will use multiple methods to build predictive models of counts of interest stated above. We will use negative binomial regression to model the counts in each week of influenza isolates detected in Allegheny county as a function of total influenza-like illness related absences reported in our observation schools, total absences, total influenza positives reported from our observation schools and cohorts, indicators of day of week and week of year (or other seasonal component). We will investigate multiple lags of each of these covariates (e.g. ILI school absences one week previous as well as ILI school absences two

weeks previous). Seasonality will be adjusted for by multiple ways in competing models including by using indicator variables of month of year, sinusoids and flexible splines to account for the effect of season over the study area.

Regression trees, Boosted Regression Trees and Lasso Regression
The number of potential models described above is large. Considering multiple covariates in addition to cause specific school absences, absences at multiple temporal lags or at multiple spatial aggregations. Formal methods are required to search the space of potential models. We will use several machine learning techniques to search the space of possible regression models to choose prediction models for each outcome that are optimal. We will employ regression trees, boosted regression trees and lasso regression to identify covariates and model structure that perform best by likelihood-based measures of model fit.

Hidden Markov Models
We will develop a statistical model of observed case counts of metrics of community wide influenza based upon Hidden Markov Models (HMMs). This model has increased flexibility to incorporate multiple types of uncertainty in the analysis and relationship of covariate data to states of the system. In this model, observations of disease incidence in discrete time periods are based upon an underlying disease incidence, and state transitions are based upon a vector of covariates.

Thresholds
Each of the methods described above for modeling count data will be used to project weekly counts of the count outcomes of interest and project the time at which particular thresholds of outcomes of interest will be exceeded. We will use receiver operator characteristic curves (ROC curves) to assess the performance of multiple models to predict the time at which particular thresholds are met.

Statistical Methods: Spatial-temporal signatures
School absences are caused by a myriad of phenomena, both infectious and non-infectious that vary by season, day of week and in response to particular events (e.g. holidays). Contagion is inherently a spatial-temporal correlated process. Absences due to influenza and other respiratory pathogens may have a unique spatial temporal signature that reflects the spatial-temporal correlation of the contagious process. We will investigate whether temporal changes occur in the correlation matrix of absences in each class i with all other classes j both within and outside of the same school. We will attempt to detect changepoints in the correlation structure that are historically associated with different parts of the influenza season. We will also estimate measures of the spatial temporal scale of clustering of transmission using strong assumptions of the transmission process (e.g. hierarchically structure transmission models) and models that do not make as strong assumptions about the transmission dynamics of 'pathogens' potentially associated with absences in each location.

Data for Influenza Surveillance
Several sources of data will be utilized to monitor for influenza infections in the region and to compare to influenza-related absences in the schools. National data of influenza most reliably comes from weekly CDC reports accessible on their website. U.S. World Health Organization

(WHO) and National Respiratory and Enteric Virus Surveillance System (NREVSS) collaborating laboratories report to CDC the number of respiratory specimens tested for influenza and the number positive by influenza virus type and influenza A virus subtype. Region specific data is available at http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html. Pennsylvania is in region three. Data for Pennsylvania ILI, influenza cases, hospitalizations and deaths are collected by the Pennsylvania Department of Health (PADOH). These data have been made available to us through a partnership between Dr. Cummings, Dr. Donald Burke and colleagues at the Pennsylvania Department of Health. Using surveillance data on influenza from the entire state we investigated spatial-temporal dynamics of influenza incidence across the state of Pennsylvania (21, 22). Weekly estimates of reported influenza cases from 2003–2009 were provided by the Pennsylvania Department of Health. Briefly, the Pennsylvania National Electronic Disease Surveillance System (PANEDSS) is a computer application used to conduct surveillance of reportable diseases including influenza. Case reports are routinely collected by providers and laboratories and are transmitted electronically to the PANEDSS system. The surveillance system defines each influenza season to begin in the 40th week of the calendar year through the last week of April of the following year.

Because regional and local data sometimes lag behind disease in the community, real-time numbers of confirmed cases will be monitored in the University of Pittsburgh Medical Center (UPMC) virology laboratory. The director of the laboratory, Dr. Charles Rinaldo, has agreed to provide reports to our study team for purposes of monitoring disease transmission and predicting incidence in schools. The UPMC laboratory serves the 15 UPMC hospitals in the region, including the Children's Hospital of Pittsburgh and more than 20 outpatient clinic sites and 8 urgent care centers. The volume of laboratory testing for ILI in 2012-13 influenza season peaked at over 800 samples per week. The laboratory reports results of tests using multiplex PCR for 18 respiratory viruses (influenza A (H3, H1); influenza B; parainfluenza 1, 2, 3; metapneumovirus; respiratory syncytial virus A and B; enteroviruses; adenovirus and rhinovirus) within 24 hours of test collection.

Reporting Results
School districts will be given a copy of reports and papers generated by this research. The investigators will make de-identified and cleaned data available to school officials.

5. Data Custody, Security and Confidentiality
Hard copies of data collection materials with identifiers will be locked in a secure cabinet or room with limited access by specific individuals.

The project will be considered to be inactive 5 years after the completion of analysis. At such time, it will be the responsibility of the principal investigator to ensure that all specimens and personal data are destroyed. Personal data will be destroyed by deleting SEFs. Biological specimens will be destroyed by incineration. When possible, redacted (de-identified) versions of the data collection sheets will be used for coding and analysis. Deidentified data will be always used for coding and anslysis.

All students will be given a study ID that will be used to link multiple pieces of data collected during the study period. Databases containing study IDs and names will be kept only at

secure computers in locked offices at the University of Pittsburgh, where study staff will access them for the purposes of conducting the necessary follow-up of students. This database will be encrypted with truecrypt file encryption. If a subject withdraws from the study, data will be deleted.

De-identified datasets, containing no personal identifiers will be created for analysis. These datasets will be transferred using secure protocols and maintain password protection using truecrypt. However, these de-identified datasets will travel outside of the University of Pittsburgh to Dr. Cummings at Johns Hopkins Bloomberg School of Public Health, where it will be kept on secure computers in locked offices and maintained in encrypted truecrypt drives. All data stored over time will be de-identified. Once the study is over, files that would identify person's name and id code will be destroyed.

All data, identifiable or not, will be stored on strongly encrypted filesystems (SEFs). This will apply to short-term data storage on notebook computers used by the field team, to USB drives used to transport identifiable personal data, to the personal computers used to store and analyze these data. Personal data will not be transmitted over the internet. The number of copies of personal data will be kept to a minimum and the project co-Is will maintain a record of all copies (and those responsible for pass phrases).

There is the possibility that confidentiality could be breached if the surveys are accessed by non-study personnel. However, we will take precautions to protect against breaches of confidentiality. No identifying information will be included in the electronic data capture methods. All paper records will be kept in locked cabinet within Pittsburgh University, and only key study personal will have access to these files. No protected Health information or other sensitive personal information will be collected; disclosure, while highly unlikely, would have no or minimal impact.

6. Recruitment Process

School Districts
School districts were recruited based on predetermined interest, identified by prior contact with the School Based Research and Practice Network. Investigators met with school district/school administration and provided a detailed summary of the research, along with the opportunity to ask questions and discuss participation.

Schools
The schools will be chosen to represent all grades and the full range of schools (elementary, middle, and high), with final selection based on recommendations from the school district leadership, and willingness to participate. The study needs schools representing each grade from K to 12.

Students and School Staff
Project staff will meet with school boards, parent-teacher organizations, school staff, school nurses, and pursue related activities in order to provide initial information about the project. A flyer (SMART2 flyer - version.9.docx) will be distributed to all parents and

school personnel. Flyers will be distributed using the mechanisms that each individual school uses to disseminate information to parents, such as through material sent home with students, school web site, mailed home, emails, etc. Parents would also be informed through personal encounters, such as PTA meetings. Subsequently, parents/guardians will be sent a concise and readable summary of the study that includes a signature section for opting-out and a full disclosure form. After reviewing this material, they can choose to opt their student out of the study by signing and returning the letter. Students can opt out at any time by simply saying that they do not wish to participate. Parents and students have access to investigators via telephone or email to answer questions, and project staff will work closely with school personnel to ensure that parents/guardians are fully aware of the project and have adequate information regarding what is involved.

Parents and guardians in households that have not opted out will be sent a flyer ("Cohort flyer – version 9.docx" included in original submission) to determine if they would like to participate. Cohort households will be selected on a first come, first served basis from these volunteers. If they would like to participate, we ask those parents to contact us to participate.

7. Consent process and documentation

Project staff will meet with school boards, parent-teacher organizations, school staff, school nurses, and pursue related activities in order to provide initial information about the project. The research team will explain the research through fliers and presentations with the goal of having all understand what is being done and why. This research presents no real risks (besides a small risk of breach of confidentiality and the discomfort of the swabbing), so deliberation of parents should be easy and should have no pressure to participate.

Parents/guardians will be sent a cover letter and concise and readable summary of the study that includes a signature section for opting-out and a full disclosure form that follows the elements for informed consent. After reviewing this material, they can choose to opt out of the study by signing and returning the letter. We are using an opt out approach in order to obtain a more representative sample of all school children. We also feel that the risk to participants is low. Students can opt out at any time by simply saying that they do not wish to participate. Parents and students have access to investigators via telephone or email to answer questions, and project staff will work closely with school personnel to ensure that parents/guardians are fully aware of the project and have adequate information regarding what is involved. If parents or students opt out, they will not be contacted when the student is absent. Data on the student's absences will still be included in the aggregate data included by schools of absences by grade, however, we have no ability to exclude these data and these will be completely de-identified. We will allow parents to opt out on initial contact over the phone, thus minimizing the burden on parents who do not wish to participate but did not opt out in writing. This will prevent children with influenza like illness from being tested upon return from an absence. There is a risk that children leaving school ill could be swabbed when a parent does not want them to be and has not opted out. We will allow the parent to opt out at anytime. We will also call the parent if children state that they are uncertain if their parents would want them to be swabbed.

In our discussions with parents, we will inform them that a letter and disclosure are being sent.

We will collaborate with the school to transmit the letter to parents, using the methods they use to communicate important information. We will use the method that the school deems best to inform parents, including first class mail.

The cohort consent form will be given to individuals that express interest. Parents will be enrolled in the cohort study to provide information on their families. Families will be recruited into the cohort study from the 6 schools participating in study that are providing individual level data. This will be an individual signed, consent with each family. Each member of the family will also include their assent on this form if under 18, with parental consent or consent if over 18.

To be compliant with FERPA regulations, school districts will not provide individual level absence information on students who have been opted out of the study, only aggregate counts.

8. Risks

This study is minimal risk for all participants involved. The probability and magnitude of harm or discomfort anticipated are not greater than those encountered in ordinary daily life or during the performance of routine physical or psychological examinations or tests.

There is minimal risk involved in completing symptom diaries. The minimal risk is of breach of confidentiality. Those participants who show symptoms of the flu or were absent due to flu-like symptoms may be asked to be tested for flu. The test involves a nasal swab to obtain a sample of mucus. There is a small risk of bleeding and discomfort from the nasal swab. Overall, the study does not interfere with everyday life of the students.

9. Benefits

There are no benefits to the participants. There may be benefits to their community and school. The primary goal of the proposed effort is to improve prediction and tracking of influenza activity in entire communities.

10. Payment

There will be small incentives for participation in school based absenteeism surveillance such as small flash drives, pens, pizza parties, etc., but the value of each item will not exceed $10/student. Each school will also receive two Apple iPads (current model) to raffle among students who participate in surveillance (regardless of being swabbed). All incentives will be distributed through the school to enhance the school mission, except, as noted below.

Families enrolling in the cohort study will be given a $20 gift card for enrolling in this study. Each week when they provide information, they will be entered into a drawing for a $200 gift card. Families will be given an additional $20 gift card if they complete at least 5 of the first 6 weekly surveys, and another $20 gift card if they complete at least 5 of 6 of the weekly surveys in weeks 7-12 of the study period. All gift cards will be sent at the conclusion of the 12 weeks.

Incentives for household cohorts will be given to head of household. Incentives for absenteeism surveillance efforts will be provided directly to students.

11. Drug Products, Vitamins, Dietary Supplements and Devices
Not applicable.

12. Safety Monitoring
A data and safety monitoring plan will be implemented by the Principal Investigator to ensure that there are no changes in the risk/benefit ratio during the course of the study and that confidentiality of research data is maintained. Each member of the study team will meet with the PI and review confidentiality issues and complete a confidentiality agreement, prior to having contact with research subjects. Investigators and study personnel will meet weekly to discuss the study (e.g. study goals and modifications of those goals; subject recruitment and retention; progress in data coding and analysis; documentation, identification of adverse events or research subject complaints; violations of confidentiality) and address any issues or concerns at that time. Minutes will be kept for these meetings and will be maintained in the study regulatory binder. Any instances of adverse events will be reported immediately to the University of Pittsburgh IRB and Johns Hopkins University IRB using standard forms and/or procedures that have been established by the IRB. The yearly IRB renewal for this study will include a summary report of the Data and Safety Monitoring Plan findings from the prior year.

13. Plan for reporting unanticipated problems/adverse events
The likelihood of an unanticipated adverse event occurring is small, as the study risks are small. If such an event occurs, staff will report this to the Project Manager, who will advise the PI, and review and document the event. The event will then be reported to the IRB using the reporting protocols/forms in place at the time of the event.

One problem that could arise is the feeling among parents that the study is taking away from the education of their children or is someway damaging to their students. The success of the study hinges on the acceptance of the study by parents. To help foster a positive relationship with parents, parents will be contacted through multiple means to inform them of the study goals and procedures. Presentations will be made or information booths present at PTA meetings, student concerts and other after school meetings that parents attend.

Another problem that we have sometimes encountered is a lack of reliable contact information for parents. We will try to circumvent this problem by requesting updated contact information at multiple times during the school year and by multiple forms of communication (email, text, phone call).

Sample handling is always a concern with viral samples as well as sampling technique by individual nurses to determine if they are performing consistent and correcting swabbing. We will periodically assess viral detection among staff at different schools to determine if any staff experience low rates of viral detection suggesting that their execution of protocols is sub-standard.

14. Other IRBs

University of Pittsburgh IRB (see attached documentation)

Contact: Christopher Ryan

ryancm@upmc.edu

Main Phone: (412) 383-1480
Main Fax: (412) 383-1508

Mailing Address:
3500 Fifth Ave.
Hieber Building, Suite 106
Pittsburgh, PA 15213

15. Outside collaborations

	JHSPH	University of Pittsburgh
Primary Grant Recipient	Derek Cummings	
Subcontractor		Shanta Zimmer

For the following, indicate "P" for "Primary", "S" for "Secondary" as appropriate to role and level of responsibility.) Add additional items if useful.

1	Human subjects research ethics training for data collectors	S	P
2	Day to day management and supervision of data collection	S	P
3	Reporting unanticipated problems to the JHSPH IRB/Sponsor	P	S
4	Hiring/supervising people obtaining informed consent and/or collecting data	S	P
5	Execution of plan for data security/protection of participant data confidentiality, as described in Sect. 5.	P	S
6	Biospecimen processing, storage, management, access, and/or future use	S	P

16. Oversight plan for student studies

None planned.

17. Oversight plan for studies conducted at non-JHSPH sites, including international venues, for which the JHSPH investigator is the responsible PI

Dr. Derek Cummings is the Principal Investigator of this project and each of the co-investigators has extensive experience working and collaborating with him on other research projects. Dr. Shanta Zimmer is currently co-Principal Investigator with Dr. Cummings on the SMART project, a CDC funded project aimed at estimating social contact rates among school-children in multiple schools in the Pittsburgh area and estimate the risk of laboratory confirmed influenza associated with different types of contacts among students. Dr. Zimmer is an infectious disease clinician who has performed studies of the community transmission and burden of influenza in the 2009 pandemic and of influenza in inter-pandemic periods. Mr. Chuck Vukotich will serve as the senior project manager of the proposed work. Mr. Vukotich has extensive experience in public health project management stemming from 30 years of work at the Allegheny County Health Department (ACHD) as Assistant Deputy Director, where he was responsible for all service and regulatory programs of the Health Department and managed over three hundred staff working in Allegheny County. Mr. Vukotich served as senior project manager for both the PIPP and SMART projects with Dr. Cummings. They have worked together for 7 years. Dr. Cummings has conducted two CDC funded influenza projects in Pittsburgh school-children as well as a community transmission study of influenza in southern China.

Mr. Vukotich will provide day to day management of on site operations at the University of Pittsburgh, with Drs. Cummings and Zimmer, providing overall guidance and direction. Mr. Vukotich will manage all data collection. Mr. Vukotich has previously served in this role for the PIPP and SMART projects. All study staff have received or will receive training on human subjects research protection, as part of the IRB process at the University of Pittsburgh. Mr. Vukotich, Dr Zimmer and Dr. Cummings are in daily contact by telephone and email. A weekly call of all study staff will provide additional coordination.

Dr. Cummings will provide oversight of field activities by conducting multiple site visits during the initial phases of the work and during the conduct of the study.

18. Creation of a biospecimen repository

Biospecimens of participants will not be stored. There will be no future research using participant cells such as producing cell lines or genomic analysis. Specimens from nasal swabs will only be used to take viral samples. Viral samples will eventually be destroyed after being sequenced.

19. Data Coordinating Center

NA