Title:

Decipher my Data - Investigating the association between school absence prevalence collected through scientific engagement with influenza surveillance data.

Authors:

RW Aldridge[1], AC Hayward[1], N Field[1], C Warren-Gash[1], C Smith[1], R Pebody[2], D Fleming[3], S McCracken[4], on behalf of the Decipher my Data project and schools.

Affiliations:

1. Research Department of Infection and Population Health, University College London.
2. Respiratory Diseases Department, Public Health England, Colindale, UK
3. Pate's Grammar School, Cheltenham, UK
4. Gallomanor Communications Ltd, Bath, UK

Abstract

Background:

School aged children are a key link in the transmission of influenza. Most cases have little or no interaction with health services and are therefore missed by the majority of existing surveillance systems. As part of a Public Engagement with Science project, this study aimed to establish an electronic system for the collection of routine school absence data and to determine if school absence prevalence was correlated with established healthcare surveillance measures for circulating influenza.

Methods:

We collected data for two influenza seasons in 2011/12 and 2012/13. The primary outcome was daily school absence prevalence (weighted to make it nationally representative) for children aged 11 to 16. School absence prevalence was triangulated graphically and through univariable linear regression to Royal College of General Practitioners (RCGP) influenza like illness (ILI) episode incidence rate, and as a counterfactual analysis, to national microbiological surveillance data on the proportion of samples positive for influenza (A+B), Rhinovirus, RSV and laboratory confirmed cases of Norovirus.

Results:

A website was created that enabled 27 schools to submit data over the two respiratory seasons. During the first season, levels of influenza measured by established surveillance were low. In the second season, a peak of school absence prevalence occurred in week 51, and week 52 in established influenza surveillance systems. Linear regression showed a strong association between the school absence prevalence and RCGP ILI, laboratory confirmed cases of influenza A & B, and some evidence for a linear association with Rhinovirus and Norovirus.

Interpretation:

This study provides some evidence for the validity of using routine school absence prevalence as a novel tool for influenza surveillance in children. It was not possible to establish conclusively whether school absence prevalence detected outbreaks of disease earlier than existing data, and future work should examine this further.

Funding:

Wellcome Trust, People Award, 096802.

Introduction

School-aged children are an important group in the transmission of influenza with well documented outbreaks of influenza and influenza like illness in schools.(1) Studies examining the community burden of seasonal and pandemic influenza have demonstrated that children have significantly higher rates of PCR-confirmed disease and serological infection with influenza A than adults.(2)

Many people with influenza do not see a doctor as it is often a mild self-limiting illness, however, the majority of surveillance systems use data collected from patients interactions with the health service.(2) Current surveillance systems therefore may be less sensitive to milder forms of the infection and underestimate the amount of influenza transmission within the population, particularly in the early stages of an epidemic and in children.

Boarding schools taking part in the Medical Officers of Schools Association scheme send reports of various illnesses, including ILI to Public Health England for surveillance purposes during school term. Most of the schools taking part in this programme are located in the South of England and the majority collect data on boys aged 13 to 18 years. Most pupils attending these schools are resident and therefore interactions between them and the community. For these reasons, such surveillance data is unlikely to be representative of most children in England.

Pilot studies using school absence data for influenza surveillance have been conducted previously in England and demonstrated the potential usefulness of such an approach. The majority of data from these studies was collected in primary schools, with three secondary schools providing data for one influenza season.(3,4) These studies relied on historic electronic records from schools or daily upload of absence data, which can be administratively burdensome with minimal benefit to schools providing this information. We sought to address these issues by receiving data from schools across England on a weekly basis and involving students in the collection, submission and analysis of data for educational purposes.

This study aimed to establish an electronic system for the collection of routine school absence data as part of a Public Engagement with Science project and to determine if school absence prevalence was correlated with established healthcare surveillance measures for circulating influenza.

Methods

Schools that had taken part in a previous scientific engagement project (I'm a Scientist, Get me out of here! or IAS Debate Kits) were invited to participate in Decipher my Data.(5) Secondary schools in England with pupils aged 11-18 were eligible. We collected data for two influenza seasons in 2011/12 and 2012/13. Schools were all located in England, and, to make the results comparable to Public Health England's (PHE) national surveillance data, South, Central and North geographical regions were used to classify the region of each one taking part.(6,7)

After consent had been received from the Head teacher, schools uploaded basic data including the number of pupils in each year group, the percentage of students in different ethnic groups, the percentage of children on free school meals and the full time equivalent number of teachers. Each week, schools were asked to submit the total number of half-day absences for medical reasons in each year group via the project website - http://flu.deciphermydata.org.uk/. Data submission was encouraged from week 38 (September) until week 13 and 12 (March) in the 2011/12 and 2012/13 seasons respectively. Data were collected for these periods as it was likely to capture the peak period of influenza circulation. A password protected website was built that enabled school teachers and pupils to submit data each week and take part in the learning activities specifically designed to engage students with data analysis. Detailed instructions were provided on how the data should be collected and uploaded, and telephone advice was provided in case difficulties were encountered. No face-to-face training was provided. Schools were asked to submit data in as timely a manner as possible, and email reminders about the submission of data were sent on an ad hoc basis.

No details were collected about the medical reason for each absence as schools do not routinely collect these. In the second year of the project, weekly school absence data were automatically time stamped when uploaded, allowing analysis of the time lag between the end of a school week and the time to submission on the project website.

The primary outcome was prevalence of daily school absence prevalence. This was calculated using the number of school absences per day as the numerator, and the number of pupils in years 7-11 as the denominator. Consistent with previous studies, we analysed data for years 7-11 only, as children in years 12 and 13 tend to have higher levels of scheduled absence, due to the more variable nature of their school timetable and provision for personal study time, making the denominator data less reliable.(3) Schools submitted the aggregate number of absences for each year group on a weekly basis, along with the number of half-day sessions at the school (e.g. 10 represented a full 5 day week). Poisson regression was used to calculate daily school absence prevalence weighted by region. Weights were generated by using the proportion of children in years 7 to 11 for each region in England using school census data for 2010, and the proportion of children for each region taking part in this project and submitting their data each week.(8) Weighting was performed to account for differences in the proportion of children sampled in each region and proportion of all children attending schools that region. In weeks where there were no data submitted by schools for a particular region, weighting was calculated using the two remaining areas.

We anticipated the recruitment of around 100 schools. This would enable a 2% daily prevalence of school absence to be calculated with 95% CIs of 1.9-2.1% (off-peak influenza) and a 7% prevalence with 95% CIs of 6.8-7.2% (peak influenza).

Descriptive analysis was performed to examine the number of schools submitting data each week, the time to upload data (i.e. lag between end of school week and receipt of data, second year of data only), the total number of pupils in the sample population and their geographical distribution.

To examine the association between weekly weighted school absence prevalence and levels of influenza circulating in the community, results were triangulated graphically and through univariable linear regression against Royal College of General Practitioners (RCGP) influenza like illness (ILI) episode incidence rate per 100,000 population for all ages, and rates among individuals aged 5-14.(9) School absence data were also plotted against microbiological surveillance data (Datamart) on the proportion of samples positive for influenza (A+B). DataMart is based on laboratory results collated from a network of 16 PHE and NHS laboratories in England and includes respiratory swabs from primary and secondary care. These swabs are tested for a variety of viruses using real time polymerase chain reaction assays.(6) Analyses were conducted for weeks 38 to 13, the periods when schools were being actively encouraged to submit their data. All RCGP and Datamart data were taken from the weekly reports and therefore may differ slightly to what is written in the final annual reports produced by HPA/PHE.

School absence prevalence data were plotted against Datamart data for respiratory syncytial virus (RSV) and Rhinovirus and laboratory confirmed cases of Norovirus reported in Public Health England's weekly health protection report.(10) These infections are common in children and might potentially explain trends in school absence prevalence as the medical reason for each school absence is not routinely collected and therefore we did not ask schools to submit these data. This counterfactual analysis was therefore conducted to examine this possible alternative explanation for any associations found by triangulating results graphically and through univariable linear regression. All analyses were conducted in Stata version 12.

This work was conducted as a public engagement in science project and several interactive lesson plans were developed for schools taking part. Topics covered in these sessions included an introduction to the data, how to analyse the results, and how to write up the results. The project team wrote regular blogs about the project that were posted on the study website and emailed to students and teachers. During the first year of the project, students were able to post questions to scientists taking part, and in both years of the project, students were able to write 'Lab logs' about their observations and analysis, which the authors responded to.

Results

Establishing an electronic system for the collection of routine school absence data

Following an introductory email in the summer term of 2011, 352 teachers expressed an interest in taking part. Of these, 83 registered with the project website in the autumn term of 2011. Across the two seasons, a total of 47 schools returned consent forms and 27 submitted data at least once. There were a greater number of schools in PHE's southern region (16/27) compared to central (8/27) and northern geographical areas (3/27).

Schools were encouraged to submit absence levels from week 38 until week 13 in the 2011/12 season and until week 12 in 2012/13 (the start of Easter holidays). Several schools provided historical data from week 36 and continued to submit data until week 27 in 2011/12 and week 19 in 2012/13. A mean of 11 (range: 3 – 17) schools submitted data each week during the 2011/12 season and 8 schools during 2012/13 (Figure 1). Data from these schools provided absence information for a mean of 10,231 pupils per week in the 2011/12 and 7,743 pupils per week in 2012/13. No data were uploaded in weeks 43 and 7 in 2011/12, as these were school half terms. In both years no data were uploaded for two weeks during the Christmas holidays. As there was variation in the timings of half term across the schools taking part during the second season, data were uploaded for all other weeks.

Determining if school absence prevalence is associated with established surveillance measures of influenza.

During 2011/12, the mean weighted daily prevalence of school absence due to illness for children in years 7 to 11 was 2.7% (95% CI 2.5, 2.9) between week 38 and 13. There was a peak of 4.1% (95%CI: 3.1, 5.4) in week 6 (Figure 2). Week 7 was a school half term and therefore no data were submitted, and levels in week 8 reduced back to 2.2% (95%CI: 1.3, 3.6). RCGP ILI data (all ages) started to increase in week 5 and peaked during week 7 at 20.1 per 100,000 during the 2011/12 season. This peak was later than seen in previous years and did not cross the threshold of 30 per 100,000 for the entire season, indicating low influenza activity. Datamart microbiological surveillance data on the proportion of samples positive for influenza

(A+B) peaked in week 8 at 17.6%, which was also lower and later than the previous two seasons.(6)

During 2012/13, the mean weighted daily school absence prevalence due to illness was 3.4% (95% CI 3,1, 3.7) between week 38 and 12. School absence prevalence peaked in week 51 at 6.8% (95%CI 5.2, 8.9). School absence prevalence reduced to mean seasonal levels in week two at 2.8% (95%CI 2.2, 3.5). RCGP clinical surveillance data of ILI (all ages) peaked at 32.7 per 100,000 in week 52 and was followed by a second smaller peak of 20.2 per 100,000 in week seven. Daily school absence prevalence also showed an increase during this second period, reaching a maximum of 3.8% in week eight (95%CI 3.3, 5.1). The percentage of samples positive for influenza (A+B) in Datamart peaked in week one at 23.1%, and also showed a second smaller peak towards the end of February and the beginning of March. (10) Schools data generally peaking a week before the national surveillance data.

Microbiological surveillance data for RSV in 2011/12 showed the highest percentage of positive cases in week 52 at 35.7% (Figure 3). In 2012/13 the peak of RSV occurred in week 49 at 34.9%. The percentage of samples positive for Rhinovirus were highest during week 40 at 25.3%. Rhinovirus levels were also highest at the beginning of the 2012/13 season (31.8% in week 41). The total number of virologically confirmed cases of Norovirus peaked at 464 in week 7 during 2011/12 and 494 in week 50 in 2012/13 (Figure 4). There was no obvious descriptive association between school absence prevalence and these non-influenza respiratory and gastrointestinal infections across the two seasons.

There was strong evidence for a positive linear association between school absence prevalence due to illness across the two seasons with RCGP ILI data for all age groups. For every 1% increase in school absence prevalence, the rate of ILI went up by 0.010 per 100,000 (95%CIs: 0.061, 0.14; p-value <0.001; R^2 0.39; Table 1). There was also strong evidence for an association with RCGP influenza like illness in children aged 5-14 years. For every 1% increase in school absence prevalence, the rate of ILI went up by 0.063; (95%CIs: 0.041, 0.085; p-value <0.001; R^2 0.25). The proportion of samples positive for influenza A + B submitted to Datamart increased by 0.057 for every 1% increase in school absence prevalence (95%CIs 0.026, 0.087; p-value <0.001). The proportion of variance explained by linear regression was greatest for RCGP ILI in children aged 5-14 years, which explaining 54% of the association. There was also some evidence for a linear association with Rhinovirus (negatively correlated) and Norovirus (Table 1).

Public engagement in science

32 Lablogs were completed in year 1 and 18 in year 2. Year 1 also had 50 questions from students that were answered by scientists. The project website was visited 13,604 times by 8,266 unique visitors over the project period. 46,750 pages were viewed. Website activity was was slightly busier in year one. Full results of an independent evaluation of the public engagement aspect of the first year of Decipher my Data are available online.(11) Schools taking part in the project were able to provide interesting additional insights and potential explanations for differences between data from their school and national school absence prevalence, such as weather, teaching environment and other characteristics about the school or local area.

Discussion

This study aimed to establish an electronic system for influenza surveillance using public engagement with science, and investigate whether school absence prevalence due to illness was correlated with established surveillance measures for influenza. The project ran during the 2011/12 and 2012/13 seasons and recruited 47 schools with data submitted by 27. In both seasons, peaks in school absence prevalence occurred one weak before peaks in national data. Linear regression showed a moderate association between school absence prevalence and RCGP reported influenza like illness, laboratory confirmed cases of influenza A & B, and some evidence for a linear association with Rhinovirus (negatively correlated) and Norovirus.

Strengths and weaknesses of the study

This study has demonstrated the feasibility of establishing a novel mechanism for influenza surveillance in schools using public engagement in science. The study not only resulted in academically interesting findings, but served an important educational purpose. Students taking part were provided with a unique opportunity to analyse data and learn about basic epidemiological concepts and the interpretation of scientific results. The fact that levels of influenza were low during the first year, whilst frustrating from a research perspective, was educationally useful, demonstrating how not everything in science goes exactly according to plan despite much preparation.

There are several limitations of using school absence data for surveillance of influenza activity in the community, particularly the fact that it is not possible to collect any data during school holidays. This was important during the 2012/13 season when traditional surveillance indicated a peak during week 52, which fell during the Christmas holidays. School absence prevalence demonstrated a peak during week 51, but it was not possible to examine levels during week 52. As a result, it was not possible to examine further whether school absence prevalence peaked earlier than other markers of influenza activity. It was not possible to analyse national surveillance data for comparable age categories to those collected by this research project. Clinical reason for an absence was not collected, and despite using only data for illness absences it is likely that other illnesses rather than flu were the cause in some cases..

Schools taking part in Decipher my Data were predominantly based in the South of England, and despite weighting results to make them nationally representative, the convenience sampling and low number of schools in the North may have lead to a bias and greater levels of uncertainty within these estimates. Schools were given detailed instructions on how to collect and upload data, however, measurement bias (due to inconsistencies in the way the data were processed and uploaded may have varied across schools taking part) is likely to have been randomly distributed and will therefore lead to a non-differential bias and a reduction in the study power.

Comparison to existing literature

A previous study conducted in England recruited eight primary schools and three secondary schools during the 2005/6 season.(3) This study was carried out for one season and included self-reported cause of the illness by the parent or guardian at the time of notifying the school about a child’s absence. The RCGP ILI peak occurred one week after school absence data. A similar study was also performed during 2005-2007 using results from six primary schools in a single local authority in East London.(4) This study was able to calculate both the incidence and prevalence of school absences from the data collected. Peaks in the prevalence of school absence showed a greater correlation with laboratory confirmed cases of influenza A & B than incidence rates.

A further study used public engagement in science to investigate the transmission of influenza in young children.(12) The study used children aged 13-15 to capture mixing patterns in children aged 4-11. Data collection questionnaires were designed in association with the school children and administered by the students aged 13-15. The study found evidence of sex-specific assortive mixing within and between classes in the same school, and a marked social structure. The authors concluded that the methods were a helpful way to examine mixing patterns in this difficult to research group.

Interpretation of the findings

The results of this study provide evidence that school absence prevalence could be a useful tool for surveillance of influenza in children aged 11 to 16 and may have utility for providing earlier warning of an outbreak than existing measures. The data are likely to be more representative of the community burden of disease in children compared to existing surveillance data. Schools were actively engaged in the collection and analysis of the data. It was not possible to establish with certainty whether school absence prevalence detected outbreaks of disease earlier than existing data, or whether peaks occurred at an earlier stage, and future work should be carried out to examine these possibilities as such data would enable a more timely public health response to influenza epidemics and pandemics.

Tables

Surveillance data	Beta coefficient (95% CIs)	R^2	p-value
RCGP influenza like illness (All ages, rate per 100,000)	0.010 (0.061, 0.14)	0.39	<0.001
RCGP influenza like illness (5-14 years, rate per 100,000)	0.063 (0.041, 0.085)	0.54	<0.001
Influenza A + B (Percentage of laboratory samples testing positive)	0.057 (0.026, 0.087)	0.25	0.001
RSV (Percentage of laboratory samples testing positive)	0.0051 (-0.021, 0.031)	0.004	0.694
Rhinovirus (Percentage of laboratory samples testing positive)	-0.073 (-0.12, -0.026)	0.19	0.003
Norovirus (Total number of laboratory samples testing positive)	0.0026 (0.00021, 0.005)	0.10	0.034

Table 1. Univariable linear regression models examining association between weighted school absence prevalence (years 7-11) and surveillance data for respiratory infections and Norovirus

Figures

Figure 1. Number of schools submitting absence data each week

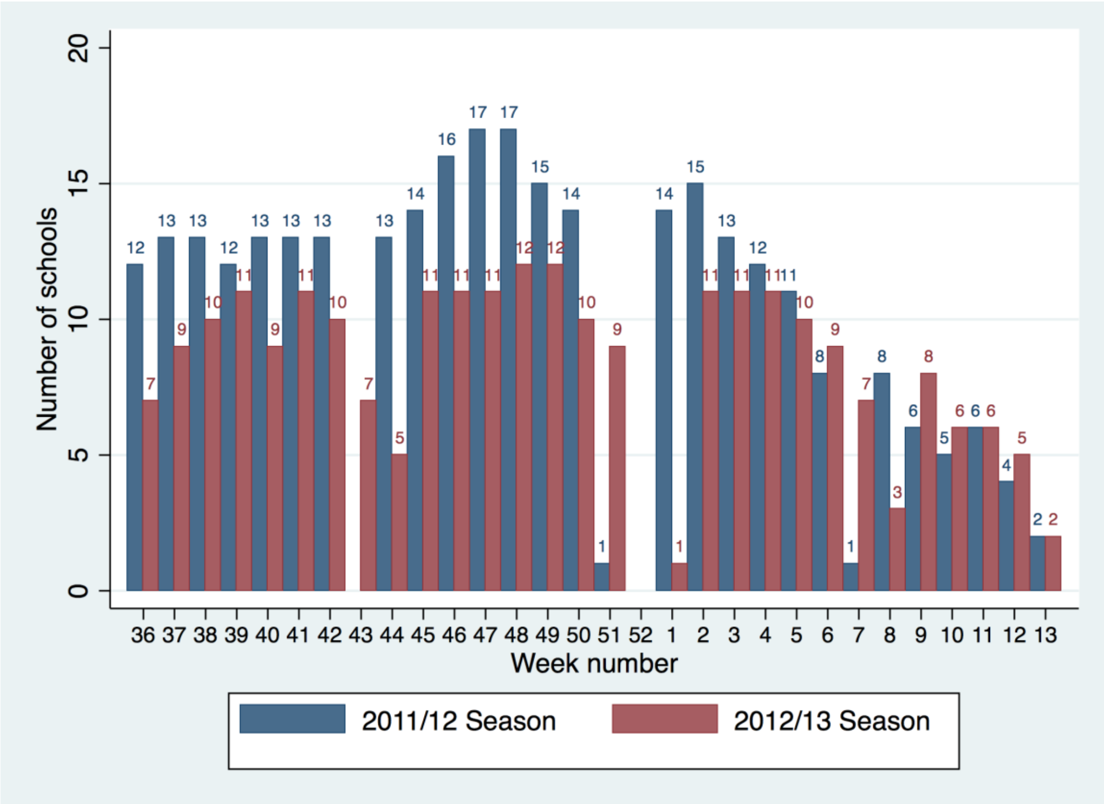


Figure 2. Daily school absence prevalence due to illness (years 7-11), weekly influenza like illness consultation rates to sentinel general practices in England, and weekly proportion of samples positive for influenza (A & B) from Datamart.

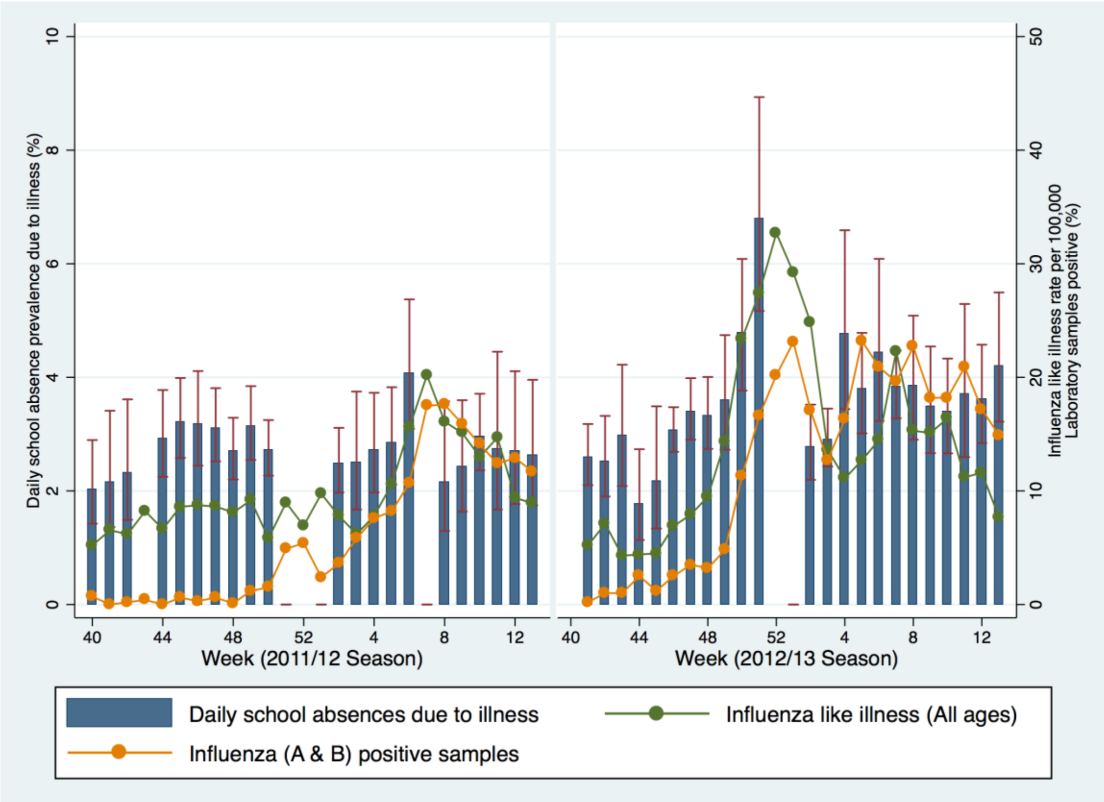


Figure 3. Daily school absence prevalence due to illness (years 7-11) and the weekly proportion of samples positive for RSV and Rhinovirus from Datamart.

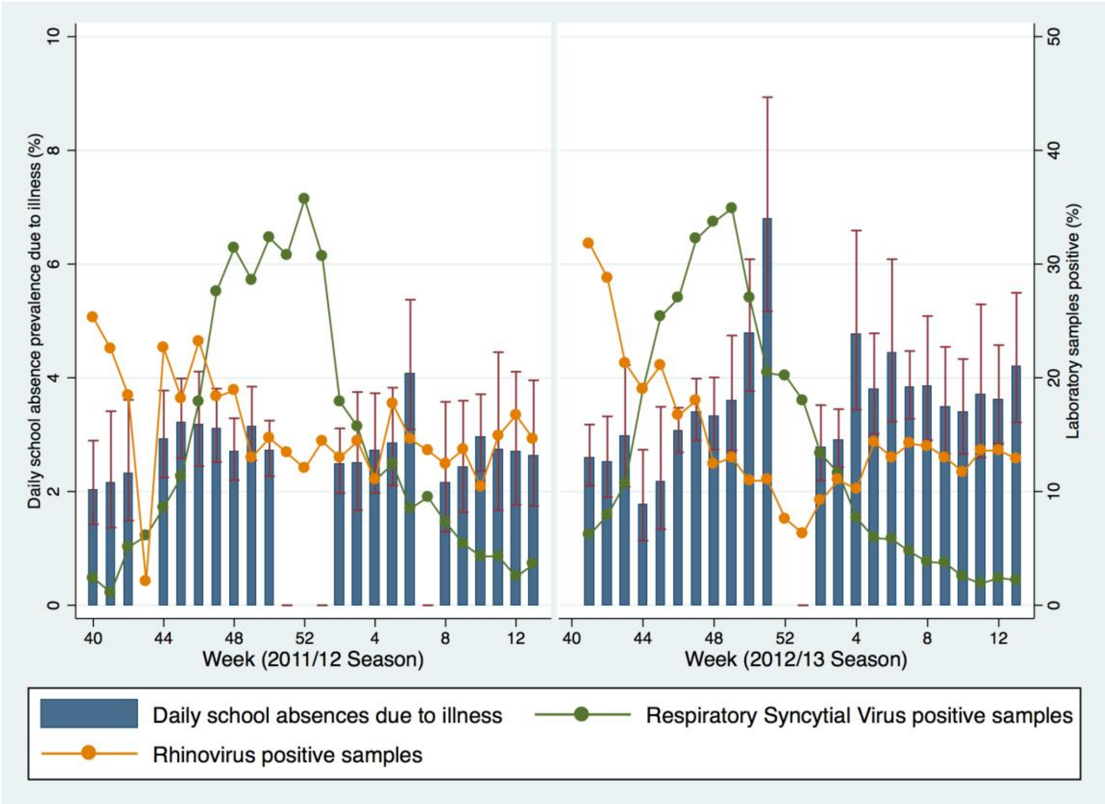
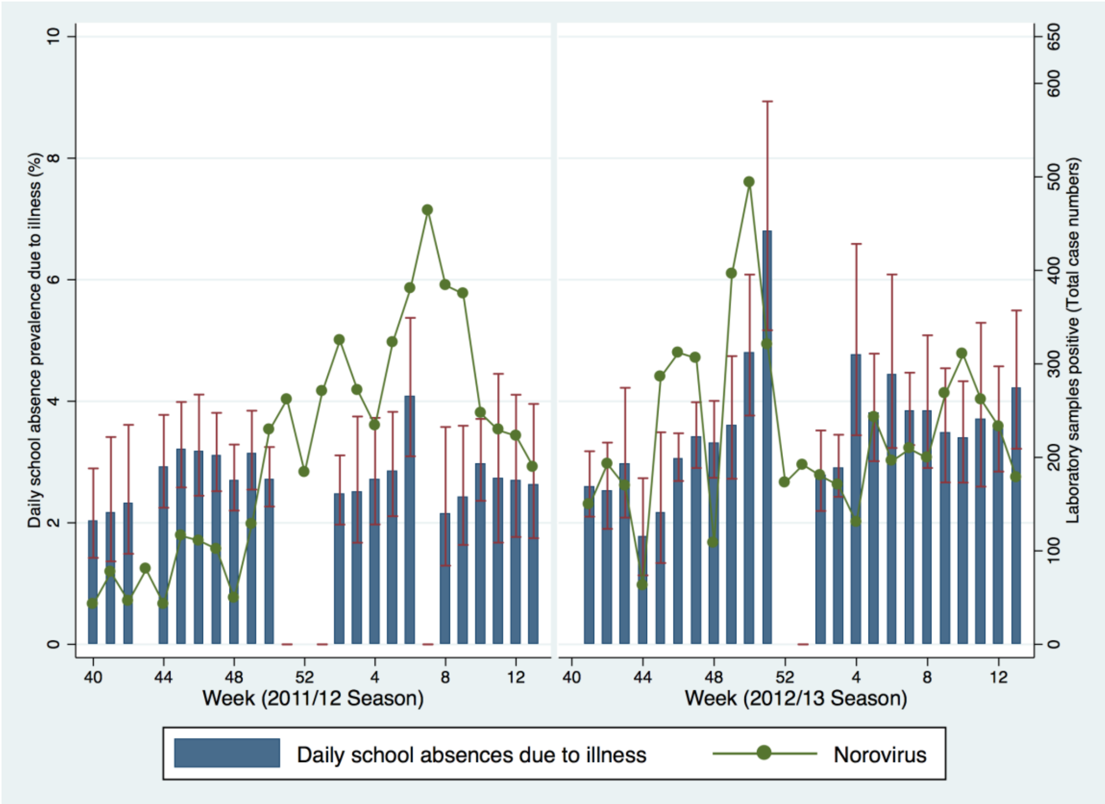


Figure 4. Daily school absence prevalence due to illness (years 7-11) and the weekly number of laboratory confirmed cases of norovirus.



References

1. Zhao H, Joseph C, Phin N. Outbreaks of influenza and influenza-like illness in schools in England and Wales, 2005/06. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2007 May;12(5):E3–4.

2. 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 [Internet]. 2014 Mar [cited 2014 Apr 7]; Available from: http://www.thelancet.com/journals/lanres/article/PIIS2213-2600(14)70034-7/abstract

3. Mook P, Joseph C, Gates P, Phin N. Pilot scheme for monitoring sickness absence in schools during the 2006/07 winter in England: can these data be used as a proxy for influenza activity? Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2007 Dec;12(12):E11–12.

4. Schmidt WP, Pebody R, Mangtani P. School absence data for influenza surveillance: a pilot study in the United Kingdom. Euro Surveill. 2010;15(3).

5. I'm a Scientist, Get me out of here! | A science outreach education and engagement activity [Internet]. [cited 2014 Mar 13]. Available from: http://imascientist.org.uk/

6. Agency HP. Surveillance of influenza and other respiratory viruses in the UK: 2011-2012 report [Internet]. [cited 2014 Mar 13]. Available from: http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1317134576275

7. Agency HP. Surveillance of influenza and other respiratory viruses, including novel respiratory viruses, in the UK: Winter 2012-13 [Internet]. [cited 2014 Mar 13]. Available from: http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1317139320524

8. Schools, pupils and their characteristics: January 2010 - Publications - GOV.UK [Internet]. [cited 2014 Mar 13]. Available from: https://www.gov.uk/government/publications/schools-pupils-and-their-characteristics-january-2010

9. Fleming DM, Crombie DL. The incidence of common infectious diseases: the weekly returns service of the Royal College of General Practitioners. Health Trends. 1985 Feb;17(1):13–6.

10. Public Health England. Health Protection Report [Internet]. [cited 2014 Mar 14]. Available from: http://www.hpa.org.uk/hpr/

11. I'm a Scientist, Decipher my data! | Real science, real data. [Internet]. [cited 2014 Mar 14]. Available from: http://deciphermydata.org.uk/

12. Conlan AJK, Eames KTD, Gage JA, von Kirchbach JC, Ross JV, Saenz RA, et al. Measuring social networks in British primary schools through scientific engagement. Proc R Soc B Biol Sci. 2010 Nov;278(1711):1467–75.