

# Timeliness of syndromic influenza surveillance through work and school absenteeism

by

Bollaerts K<sup>1</sup>, Antoine J<sup>1</sup>, Robesyn E<sup>2</sup>, Van Proeyen L<sup>3</sup>, Vomberg J<sup>4</sup>,  
Feys E<sup>5</sup>, De Decker E<sup>6</sup>, Catry B<sup>1</sup>

---

## Abstract

*In this paper, we investigate the usefulness of work and school absenteeism surveillance as an early warning system for influenza. In particular, time trends in daily absenteeism rates collected during the A(H1N1)2009 pandemic are compared with weekly incidence rates of influenza-like illness (ILI) obtained from the Belgian Sentinel General Practitioner (SGP) network. The results indicate a rise in absenteeism rates prior to the onset of the influenza epidemic, suggesting that absenteeism surveillance is a promising tool for early warning of influenza epidemics. To convincingly conclude on the usefulness of absenteeism data for early warning, additional data covering several influenza seasons is needed.*

## Keywords

*School absenteeism, worker absenteeism, influenza, influenza A virus, H1N1 subtype*

## Introduction

Early warning surveillance systems for influenza, and for emerging infectious diseases in general, are indispensable tools for (pandemic) preparedness. Evidently, control strategies like antiviral therapy campaigns, vaccination campaigns and business continuity plans are most effective when timely implemented, preferably some weeks before the onset of the epidemic. To enhance early detection of epidemics, 'syndromic surveillance' systems are used worldwide. These systems rely on existing, non-specific, prediagnostic healthcare data such as data on ambulance dispatches, pharmacy sales and absenteeism.

A review of several data sources used for syndromic influenza surveillance is given by Dailey et al. (1), discussing the strengths and weaknesses of the different data sources. Elaborated examples on the use of syndromic surveillance for early warning using school absenteeism data and ambulance dispatch data are given in Besculides et al. (2) and Bork et al. (3), re-

---

<sup>1</sup> Scientific Institute of Public Health, Directorate Public Health and Surveillance, Brussels, Belgium

<sup>2</sup> Flemish Agency for Care and Health, Department of Public Health Surveillance, Brussels, Belgium

<sup>3</sup> Flemish Agency for Educational Services, Brussels, Belgium

<sup>4</sup> Department Education, Vocational Training and Employment, Ministry of the German-speaking Community of Belgium, Eupen, Belgium

<sup>5</sup> SNCB (Belgian railways), Medical service, Brussels, Belgium

<sup>6</sup> FPS Health, Food Chain Safety and Environment, Medical expertise (MEDEX), Brussels, Belgium

spectively. One of the most important characteristics of a syndromic surveillance system is its timeliness, or the difference between the time of occurrence of a (syndromic) event and the time of occurrence of the reference standard. The latter is commonly chosen to be the influenza-like illness incidence as monitored through the SGP networks in Europe (4).

In this article, we describe the work and school absenteeism surveillance systems launched in Belgium during the A(H1N1)2009 pandemic and we explore their utility as an early indicator of influenza epidemics in Belgium. In particular, we will compare daily absenteeism rates with weekly incidence rates of influenza-like illness (ILI) obtained from the Belgian GPs network (5).

## **Material and methods**

### **Worker absenteeism data**

Daily worker absenteeism data were obtained from the Belgian Medical Expertise (MEDEX) on a daily basis and from the Belgian railway (SNCB/NMBS) twice a week. For MEDEX, data concerned the worker absence of the Federal Public Service Health, Food Chain Safety and Environment representing +/- 1650 workers. Data were collected for all weekdays starting from 19 August 2009. The variables were the daily number and percentage of employees being absent as well as the number of newly declared absent employees each working day.

For SNCB/NMBS, the daily number of persons reported ill was recorded for the period starting from 1 July 2009 (weekends and holidays included). The total number of active personnel was calculated once each month (+/- 38,700 workers).

### **School absenteeism data**

The collection of school absenteeism data was organised by the three Belgian authorities competent for education, representing students (from 3 to 18 years of age) within the Flemish Community ( $\pm$  1,110,000 students), the French Community ( $\pm$  867,000 students) and the German-speaking Community ( $\pm$  13,700 students). All communities reported on a weekly basis. For the Flemish Community, data were collected through a web-based application starting from 1 October 2009, based on a systematic randomised sample of 307 schools, representing 5% of all Flemish schools, splitting them in nursery (3-6 years), primary (7-12 years) and secondary schools (13-18 years). Representativeness at the Flemish level was pursued by stratifying on the above-mentioned three age groups, on geographical distribution over five provinces, on number of students per school and on degree of urbanisation (large city, small city and rural area). Registration was web-based, using a school-specific password. The variables were the daily number of registered students and the daily number of absences, due to illness, motivated by general practitioners or, at the condition that the period of absence is at most 3 days and not more than 4 times per school year, motivated by parents. For nursery schools, overall absenteeism was registered, because no mandatory schooling exists in this age group and proof of illness has not to be declared. Half-day ab-

sences were not recorded. Of the sample of 307 schools, a median of 272 schools (89%) (minimum 230, maximum 279) registered absenteeism between 1 October 2009 and the start of the Christmas holidays on 18 December 2009.

For the French Community, data were collected starting on 5 October 2009 through a web-based recording system, through which all schools of the French Community were invited to report the daily number of registered students and the daily number of student absences. The participation rate was 40%. Schools were subdivided into basic education (nursery and primary schools), and secondary education.

Finally, for the German-speaking Community, data were collected starting from 7 September 2009. All schools from the German-speaking Community were requested to weekly report the daily number of student absences. The number of registered students is set twice each academic year, namely on 30 September and on 30 January. The absenteeism rates for September and October 2009 were calculated using the number of students registered on 30 January 2009, whereas the remaining absenteeism rates were calculated using the number of students registered on 30 September 2009. Schools were subdivided into nursery and primary schools, for which the participation rate was 75%, and secondary schools, for which the participation rate was close to 100%.

#### Influenza-like illness (ILI) incidence

The reference standard used to assess the timeliness of the different absenteeism data sources is the weekly ILI incidence (per 100,000 inhabitants), derived from the Belgian GPs influenza surveillance database of the SGP network, coordinated by the WIV-ISP (Scientific Institute of Public Health). Detailed information on this network can be found elsewhere (5). The ILI incidence plays an important role in declaring an influenza epidemic in Belgium. In particular, one of the three criteria to be met to declare an epidemic is an ILI incidence above the threshold of 141 ILI cases per 100,000 inhabitants (5). In 2009, the onset of the A(H1N1)2009 epidemic in Belgium was declared one week after the ILI incidence exceeded the threshold at the beginning of October (week 40).

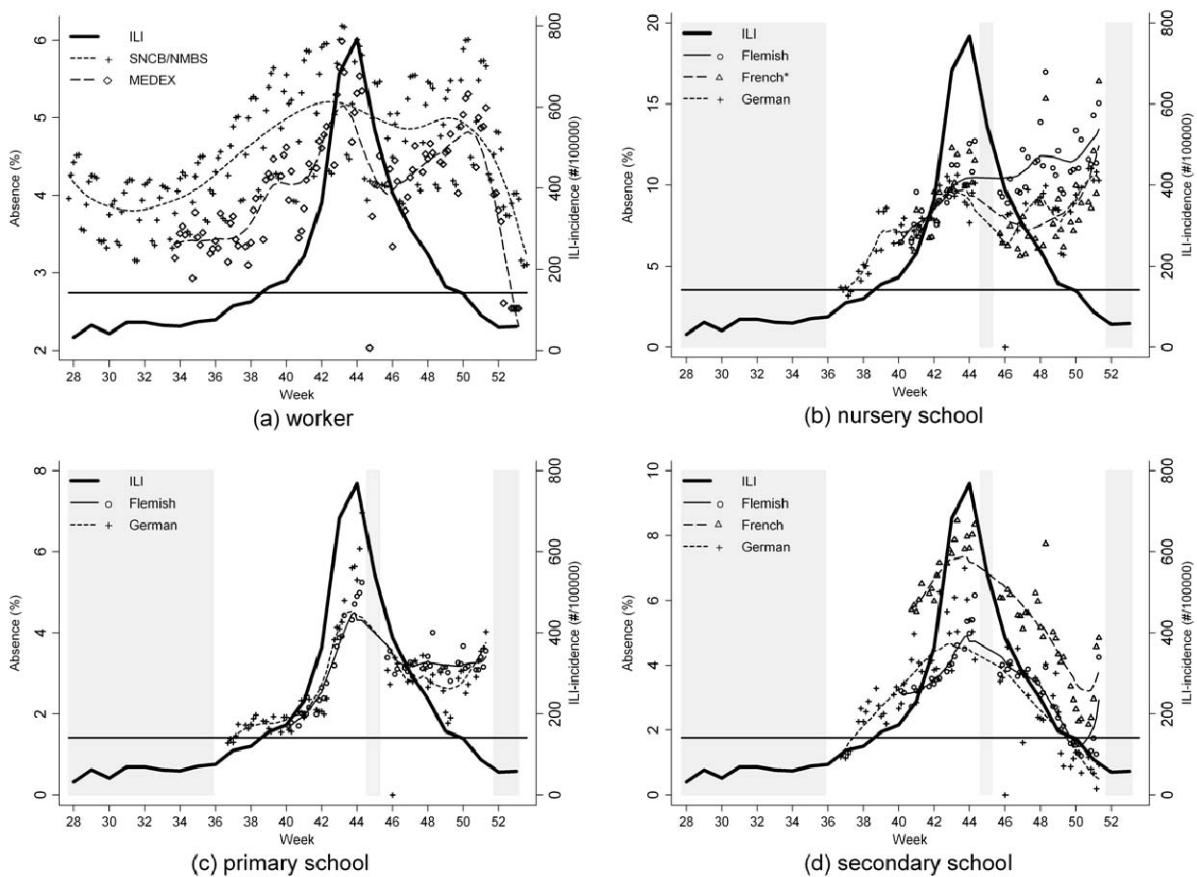
#### Statistical methods

For the two worker and three school absenteeism datasets, the daily absenteeism rates (expressed as %) were calculated starting from the beginning of the data collection until the end of 2009. To visualise time trends of the absenteeism rates, locally weighted scatter plot smoothing (LOWESS) was used. The optimal bandwidth, which lies in the interval [0%, 100%], was selected using cross-validation from a set of candidate values  $\lambda = [0.1, 0.2, \dots, 1]$ . Then, the timeliness of the different data sources was explored by visual inspection, comparing the LOWESS smooths with the ILI curve. The analyses were carried out using R 2.3.1 and Stata (version 10).

**Results**

The daily absenteeism rates (%) for the different worker and school absenteeism data sources are represented by means of scatter plots in Figure 1. In this figure, the corresponding LOWESS smooths and the ILI curve are represented as well. A rise in worker absenteeism rates was observed starting from  $\pm$  week 37 and  $\pm$  week 38 for the SNCB/NMBS data and the MEDEX data, which is  $\pm$  3 weeks and  $\pm$  2 weeks prior to the onset of the A(H1N1)2009 epidemic (week 40), respectively. The peaks of the absenteeism curves preceded the epidemic peak at week 44 with  $\pm$  2 weeks. Finally, we observe the strong drop in absenteeism rates at the end of December, which is explained by the Christmas holidays (weeks 52 and 53) and the denominator of the worker absenteeism rates (i.e. the total number of active personnel) that was not adjusted for holiday absenteeism.

Figure 1. Curve of the weekly influenza-like-illness (ILI) incidence per 100,000 inhabitants (right axis) and scatter plots with smoothed curves of the daily absenteeism rates (left axis) for (a) SNCB/NMBS and MEDEX worker data, (b) nursery school, (c) primary school and (d) secondary school data by community (Flemish, French and German-speaking) during the A(H1N1)2009 epidemic in Belgium.



The straight horizontal lines indicate the ILI threshold and the gray areas indicate school holiday periods.  
 \* For the French Community, data for nursery and primary school are not available separately and are only presented in graph (b)

The collection of the school absenteeism data started only some weeks before (German-speaking Community) or even after the onset of the epidemic (Flemish and French Community). Nonetheless, some interesting observations can still be made. Although the main interest is in predicting the onset of the influenza epidemic, it is promising to see that the peaks of the primary and secondary school absenteeism rates preceded the ILI peak with  $\pm 2$  weeks. This finding is only partially confirmed by the nursery school absenteeism rates. For the German-speaking and French Communities, a first peak in nursery absenteeism was observed  $\pm 1.5$  weeks prior to the ILI peak and a second (strong) rise was observed from the beginning of December until the start of the Christmas holidays. The latter rise might be explained by a strong increase in circulation of the Respiratory Syncytial Virus (RSV), which affects predominantly young children. Indeed, the Sentinel Laboratory network coordinated by the WIV-ISP, confirmed a strong rise in RSV-positive samples since week 41 with a peak at week 50. For the Flemish Community, a rise in nursery absenteeism rates was observed from the beginning of the data collection until the start of the Christmas holidays.

Finally, we observe that both the worker and school absenteeism rates should not be used to compare absenteeism between different companies/school communities as a result of the different criteria used to register workers/students as present or absent. In addition, for the different school communities, differences in representativeness of the population under study as a result of differences in sampling methods and participation degrees adds to the lack of commensurability.

### **Discussion and conclusion**

Visual inspection suggests that worker and school absenteeism data might be used as an early indicator of influenza epidemics in Belgium. For the two worker absenteeism data sources investigated, a rise in absenteeism rates preceded the onset of the influenza epidemic by 2 to 3 weeks. The usefulness of the school absenteeism data in predicting the onset of an epidemic with the available data is harder to explore, because the recording started only a few weeks before or even after the onset of the influenza epidemic, which was partially due to the summer holidays. Nonetheless, it is promising to see that the peaks of the primary and secondary school absenteeism rates preceded the ILI peak, suggesting that school absenteeism data can also be used to predict the onset of an influenza epidemic.

Since the preliminary results are promising, it seems worthwhile to collect absenteeism data on longer time periods covering several influenza epidemics. Once such data are available, early warning algorithms can be developed taking into account e.g. seasonality, day-of-week effect, holiday effect and spatial variability. Evidently, the accuracy (i.e. the false positive and false negative rate) of such an early warning system needs to be scrutinised before implementation. In particular, high false positive rates are to be expected because several causes of increased absenteeism rates exist, especially for the younger age groups. However, by developing an early warning system that uses different data sources simultaneously, the accuracy is expected to improve substantially. Nevertheless, developing early warning sys-

tems based on absenteeism data has several advantages as well. First, because absenteeism data (like other 'syndromic' data sources) are not disease-specific, they can be used to detect a variety of circulating diseases other than influenza. Second, absenteeism data (like other 'syndromic' data sources) are prediagnostic, implying a gain in time compared to diagnostic data sources (e.g. laboratory-confirmed influenza cases). Third, and in contrast to e.g. 'syndromic' surveillance using hospitalisation or mortality data, absenteeism surveillance can potentially detect outbreaks of diseases causing minor illnesses. A fourth advantage of (school) absenteeism surveillance is that it allows monitoring age-specific subpopulations. Indeed, by monitoring nursery school absenteeism rates, a strong increase in RSV-circulation was detected, which would likely have remained unnoticed when using non-age-specific data sources. Finally, because absenteeism surveillance (like other 'syndromic' data sources) makes use of existing data, it is relatively economically beneficial and easy to collect the data. In particular, for the worker absenteeism surveillance, existing data files are sent to and analysed by the WIV-ISP. For the school absenteeism surveillance, schools summarise the collected data, report the summary data electronically in a database hosted by the corresponding authorities. Thereafter the data, aggregated over schools, are sent for analysis and reporting to the WIV-ISP.

To conclude, we would like to mention the usefulness of absenteeism data for purposes other than early warning of new epidemics. Indeed, the absenteeism surveillance system is of particular importance in providing essential information to policy makers during crisis management. For instance, absenteeism data can serve as a basis to decide whether to close down schools (locally or nationwide) as an epidemic control measure. In addition, absenteeism data are important sources to adequately assess the socio-economic burden of influenza epidemics or other disease outbreaks. However, to this end, methodological improvements (e.g. standardisation of the school/company selection criteria, standardisation of the criteria used to register workers/students as present/absent) are needed.

## References

1. Dailey L, Watkins RE, Plant AJ. Timeliness of data sources used for influenza surveillance. *J Am Med Inform Assoc* 2007 Sep;14(5):626-31
2. 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
3. Bork KH, Klein BM, Molbak K, Trautner S, Pedersen UB, Heegaard E. Surveillance of ambulance dispatch data as a tool for early warning. *Euro Surveill* 2006;11(12):229-33
4. Arkema JM, Meijer A, Meerhoff TJ, Van D, V, Paget WJ. Epidemiological and virological assessment of influenza activity in Europe, during the 2006-2007 winter. *Euro Surveill* 2008 Aug 21;13(34)
5. Van Casteren V, Mertens K, Antoine J, Wanyama S, Thomas I, Bossuyt N. Clinical influenza surveillance of Influenza A(H1N1)2009 pandemic through the network of Sentinel General Practitioners. *Arch Public Health* 2010; 68(2):62-7