Local implementation of a syndromic influenza surveillance system using emergency department data in Santander, Spain

S. Schrell¹, A. Ziemann¹, L. García-Castrillo Riesgo², N. Rosenkötter¹, J. Llorca²,³, D. Popa¹, T. Krafft¹, on Behalf of the SIDARTHa Project Consortium

¹Department of International Health, CAPHRI School of Public Health and Primary Care, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, MD 6200, The Netherlands
²Departamento Ciencias Medicas y Quirurgicas, Facultad de Medicina, Universidad de Cantabria, Santander 39011, Spain
³CIBERESP Centro de Investigación Biomédica en Red, Epidemiología y Salud Pública and IFIMA V Instituto de Formación e Investigación Marqués de Valdecilla, Santander 39008, Spain

Address correspondence to Alexandra Ziemann, E-mail: alexandra.ziemann@maastrichtuniversity.nl

ABSTRACT

Background We assessed the local implementation of syndromic surveillance (SyS) as part of the European project ‘System for Information on, Detection and Analysis of Risks and Threats to Health’ in Santander, Spain.

Methods We applied a cumulative sum algorithm on emergency department (ED) chief complaints for influenza-like illness in the seasons 2010–11 and 2011–12. We fine tuned the algorithm using a receiver operating characteristic analysis to identify the optimal trade-off of sensitivity and specificity and defined alert criteria. We assessed the timeliness of the SyS system to detect the onset of the influenza season.

Results The ED data correlated with the sentinel data. With the best algorithm settings we achieved 70/63% sensitivity and 89/95% specificity for 2010–11/2011–12. At least 2 consecutive days of signals defined an alert. In 2010–11 the SyS system alerted 1 week before the sentinel system and in 2011–12 in the same week. The data from the ED is available on a daily basis providing an advantage in timeliness compared with the weekly sentinel data.

Conclusions ED-based SyS in Santander complements sentinel influenza surveillance by providing timely information. Local fine tuning and definition of alert criteria are recommended to enhance validity.

Keywords communicable diseases, emergency care, epidemiology

Background

The European Centre for Disease Prevention and Control (ECDC) estimates that ~40 000 premature deaths occur per year in the European Union due to seasonal influenza.¹ A comprehensive and effective seasonal influenza surveillance approach can support timely and adequate reactions of professionals and decision-makers during pandemic periods.² Influenza surveillance should integrate timely surveillance and detection of outbreaks based on different data sources of different specificity as the ECDC suggests in their epidemic intelligence framework. Part of this concept is the real-time data collection from unspecific information sources known as syndromic surveillance (SyS).³ The SyS uses data collected prior to laboratory confirmation with the advantage of providing timelier information on public health threats or sole information on potentially health-threatening events for which no other surveillance information exists.³ SyS understood in such a broad sense can be indicator based (using structured data such as health-care information indicating a certain health condition or disease) and event

S. Schrell, Research Assistant
A. Ziemann, Research Associate
L. García-Castrillo Riesgo, Director of Emergency Department at Hospital Universitario Marqués de Valdecilla
N. Rosenkötter, Research Associate
J. Llorca, Professor of Epidemiology
D. Popa, Lecturer in Public Health
T. Krafft, Associate Professor of International Health

© The Author 2013, Published by Oxford University Press on behalf of Faculty of Public Health. All rights reserved
based (using unstructured information such as media reports to detect or monitor an event).\textsuperscript{5}

While there are many initiatives at local, regional and national levels, no European approach to indicator-based SyS exists. The European co-funded project ‘System for Information on Detection and Analysis of Risks and Threats to Health’ (SIDARTHa) aimed at creating the basis for an emergency medical care data-based SyS approach, which is applicable across Europe. SIDARTHa focuses on emergency data from three sources, pre-hospital emergency medical dispatch centre call logs, ambulance service run-sheets (patient records) and emergency department (ED) patient records. The project followed a local approach to alert and inform local/regional authorities about communicable and non-communicable health threats that are not at all or not timely monitored by existing surveillance systems.\textsuperscript{6}

This paper presents the results of a case-based assessment of the local implementation of the SIDARTHa SyS approach at the ED of the University Hospital Marqués de Valdecilla (HUMV) in Santander, Spain. We evaluated the system performance in terms of timeliness, sensitivity and specificity for seasonal influenza surveillance.

Methods

We conducted a retrospective quantitative analysis of two influenza seasons in 2010–11 and 2011–12 with ED data on influenza-like illness (ILI) from the HUMV in Santander, Spain.

Data sets

With an assigned population catchment of around 300 000 inhabitants, the HUMV is the largest hospital serving the Autonomous Region of Cantabria, Spain.\textsuperscript{7} The SyS system at the HUMV analyses syndromes based on patient’s chief complaint codes which are similar to the Canadian Triage and Acuity Scale.\textsuperscript{8} The chief complaints are assigned only by an emergency physician after a first examination of the patient in the ED. The ILI chief complaint code was defined in accordance with the definition of sentinel general practitioners (GP) in the region who apply the Spanish ILI definition that follows the recommendation of the ECDC:\textsuperscript{9} ‘Sudden onset of symptoms AND at least one of the following four systemic symptoms: fever or feverishness, malaise, headache, myalgia AND at least one of the following three respiratory symptoms: cough, sore throat, shortness of breath’. The Spanish authorities added to this definition the point that these symptoms occur without suspicion of any other illness.\textsuperscript{10} The data set also includes information on age, sex, severity and postal code referring to the patient’s residence. The ED data are aggregated daily and automatically uploaded in the SyS system data base during the early hours of the following day. We analysed the ED data on ILI that was available for the period 1 July 2010–30 April 2012.

We compared the syndromic data from the ED with the data from the official influenza surveillance system. The reference data set contained information on the weekly number of ILI cases as reported by sentinel general practices for the Autonomous Region of Cantabria. In Cantabria there are 17 sentinel practices who are voluntarily taking part in the national influenza surveillance system, covering 5.26% or 30 198 inhabitants of the Cantabrian population.\textsuperscript{10} The sentinel practices were chosen to represent the whole population. No information was available on the population structure of the sentinel ILI patients. ILI cases are defined only by the GP in the sentinel practices. Data were available for Weeks 44/2010–9/2011 and Weeks 44/2011–12/2012. We selected the ED data from the same periods for the comparative analysis. The sentinel system in Cantabria defines a weekly case number of >74 per 100 000 inhabitants as threshold for high influenza activity.\textsuperscript{10}

Data analysis

In a first step, we plotted the time series of the daily and weekly number of ED ILI cases against the reported weekly case numbers of the sentinel system. As influenza is a rare disease in the ED setting, normal distribution cannot be assumed. We therefore chose a one-sided cumulative sum (CUSUM) algorithm for Poisson distribution as the most suitable aberration detection algorithm for this study.\textsuperscript{11} We added the fast initial response technique to assure that large chart values do not inflate subsequent ones and control for over-production of out-of-control signals. With this technique a head start of the algorithm is obtained, which tends to give quicker signals.\textsuperscript{12} The threshold value \( b \) for the CUSUM algorithm was retrieved by a look-up procedure in the table of Lucas.\textsuperscript{11} All CUSUM calculations were done with Microsoft Excel 2003.

We performed a fine tuning of the CUSUM algorithm by testing different baselines and accepted mean values to obtain optimal surveillance results. Four different scenarios (A, B, C and D) have been tested with a receiver operating characteristic (ROC) analysis to investigate which scenario would have the best trade-off of sensitivity and specificity and to reduce the number of false-positive signals. We tested if it is suitable to take a baseline mean close to the average of influenza cases outside the influenza season, which is
anticipated to be 0 (scenarios A and B) or rather close to the number of cases occurring outside of the period of high influenza activity (scenarios C and D). Further, we tested two different accepted means, one close to the baseline mean (scenarios A and C) and one reflecting the average daily case occurrence during the influenza season (scenarios B and D). Table 1 lists the variables used for the four scenarios in relation to the different periods analysed in our study and the respective means. The ROC calculations were done using IBM SPSS 15.0.

The specificity and sensitivity of the CUSUM algorithm were calculated using a daily approach. We compared the signals of the algorithm to the weeks of high influenza activity as reported by the sentinel system. When the threshold rate of the sentinels was exceeded these weeks were counted as 7 days of expected signals by the CUSUM algorithm. Based on the signal pattern of the CUSUM algorithm we explored criteria for alerts of the SyS system that indicate the onset and end of the high activity influenza period.

Results

In the two monitored influenza seasons, around 350 ILI cases were counted at the ED per year. During the periods of high influenza activity four to five cases occurred on average per day (around 2% of all ED cases), while the daily mean during the entire year was 1.1. The majority of ILI patients were middle-aged patients with rather light symptoms. Therefore, no stratification for age or severity was carried out in the syndromic aberration detection analysis.

In both influenza seasons 8 weeks of high influenza activity were identified by the sentinel GPs (Weeks 1–8/2011 and 3–10/2012). On average, 54 cases per week occurred during the periods of high influenza activity in the sentinel system, which is equivalent to 223 cases per 100,000 inhabitants. Table 1 summarizes the main characteristics of the ED and sentinel data sets for the different periods used in our study.

The ED data corresponded to the epidemiological curve of reported sentinel cases in both seasons. While the pattern of the epidemiological curve in the season 2011–12 is comparable for the ED and sentinel system data sets, in the season 2010–11 the influenza wave was recognized earlier in the ED than in the sentinel practices (Fig. 1). In the season 2011–12 the ED case numbers were generally lower than the sentinel case numbers. In the 2010–11 season the weekly numbers of ED ILI cases were higher than the sentinel case numbers in Week 52–2. More ILI patients were

### Table 1 Properties of data sets and corresponding variables of fine-tuning scenarios

<table>
<thead>
<tr>
<th>Period</th>
<th>Total</th>
<th>Mean</th>
<th>95% confidence interval</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Standard deviation</th>
<th>CUSUM scenario variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis perioda</td>
<td>751</td>
<td>1.12</td>
<td>0.95–1.30</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>2.31</td>
<td>Accepted mean = 1 (scenarios A and C)</td>
</tr>
<tr>
<td>Influenza seasonb</td>
<td>698</td>
<td>1.91</td>
<td>1.62–2.21</td>
<td>1</td>
<td>0</td>
<td>19</td>
<td>2.87</td>
<td>Accepted mean = 2 (scenarios B and D)</td>
</tr>
<tr>
<td>High influenza activity periodc</td>
<td>492</td>
<td>4.39</td>
<td>3.70–5.08</td>
<td>4</td>
<td>0</td>
<td>19</td>
<td>3.67</td>
<td>Baseline mean = 0.0005 (scenarios A and B)</td>
</tr>
<tr>
<td>Period outside influenza seasond</td>
<td>53</td>
<td>0.17</td>
<td>0.12–0.23</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.46</td>
<td>Baseline mean = 0.5 (scenarios C and D)</td>
</tr>
<tr>
<td>Period outside high influenza activity periode</td>
<td>259</td>
<td>0.46</td>
<td>0.38–0.55</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Weekly analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>633</td>
<td>16.23</td>
<td>10.52–21.94</td>
<td>11</td>
<td>0</td>
<td>74</td>
<td>17.61</td>
<td></td>
</tr>
<tr>
<td>Sentinel GPs</td>
<td>1003</td>
<td>25.72</td>
<td>16.19–35.24</td>
<td>12</td>
<td>0</td>
<td>105</td>
<td>29.38</td>
<td></td>
</tr>
</tbody>
</table>

\( ^a \) 1 July 2010–30 April 2012.

\( ^b \) 1 October–31 March 2010–11 + 1 October–31 March 2011–12.


\( ^d \) 1 July–30 September 2010 + 1 April–30 September 2011 + 1 April–30 April 2012.


visiting the ED in the season 2010–11 (Week 44–9: \( n = 360 \)) than in the season 2011–12 (Week 44-12: \( n = 283 \)).

**Fine tuning, specificity and sensitivity**
The fine-tuning analysis showed the trade-off between the sensitivity and specificity of the different parameters of the CUSUM algorithm indicated by the area under the ROC curve (Table 2). In season 2010–11, the most suitable scenario is C with a sensitivity of 70% and a specificity of 89%. In the season 2011–12 the optimal trade-off was achieved also by scenario C with a 63% sensitivity and 95% specificity (Table 2). The variables of scenario C were applied in the case study with a threshold value \( h \) set to 4.

**Alert criteria**
On average there have been five signals of the CUSUM algorithm per week during the periods of high influenza activity. Before and after these periods only sporadic signals occurred, none with consecutive days of signals (Fig. 1). Based on this pattern we defined the criterion for alert of the SyS system for ILI as at least 2 consecutive days of CUSUM signals.

**Timeliness**
Applying the alert criterion the SyS system produced the first alert in the season 2010–11 1 week before the sentinel system reached the threshold for high influenza activity.
In 2011–12 the first alert is produced in the same week in which the sentinels reached the threshold (two signals on 2 consecutive days in Week 3) (Fig. 1). Compared with the weekly reported sentinel data timelier information is reported by the ED data-based system which reports daily.

**Discussion**

**Main findings of this study**

Daily SyS of ILI based on the ED data in Santander, Spain correlated with the outbreak pattern reported by the regional sentinel system.

A one-sided CUSUM algorithm for Poisson distributed data with the parameters of 0.5 for the baseline and 1 as the accepted mean was identified as best setting for the local analysis. We defined a local alert criterion of at least 2 consecutive days of CUSUM signals for detecting the onset and the end of the influenza season.

The SyS system indicated the onset of the period of high influenza activity in one season 1 week before and in the other season at the same time as the sentinel system. The daily reported ED data provide an earlier warning compared with the weekly sentinel reporting.

**What is already known on this topic?**

SyS is often used to detect influenza especially based on the ED data. Buehler et al. describe ED data-based surveillance as ‘common extension’ to the surveillance of seasonal influenza based on sentinel GPs. Review results suggest general usefulness of SyS for timely and sensitive information of outbreaks affecting large parts of the population such as influenza. But the performance of the systems varies depending on, for example, the aberration detection algorithm applied. As Griffin et al. suggest, fine-tuning of algorithms can be useful to enhance the performance of a SyS system.

**What this study adds**

The Poisson CUSUM algorithm was useful to detect seasonal influenza in Santander, Spain, in combination with fine tuning of parameters and definition of an alert criterion. The alert criterion reduced the number of false alerts based on single CUSUM signals and supported the timely detection of the onset and end of the high influenza activity period. Some other SyS systems apply the same alert criterion of consecutive days of signals. Further criteria explored by these SyS systems are the magnitude of the aberration, the age and severity of cases, and the activity of other syndromes and other SyS system sites. The magnitude of the aberration would not have added precision in terms of defining alerts in our case study. No significant differences in age and severity were found in our ED ILI data set, which might be due to the relatively small sample size. The criterion of signals in other syndrome groups could be a future option. Especially, the comparison to the syndrome group of respiratory illness that is established in the SyS system in Santander could be of value. Also the criterion of signal activity in neighbouring SyS sites could be explored in the future if further SyS systems would be implemented in Northern Spain. Our alert criterion needs to be assessed in the everyday work of the SyS system in future influenza seasons.

Our daily approach of calculating sensitivity and specificity provides a more precise picture of the performance of the SyS system. Other studies aggregate daily syndromic data to weeks for comparison to weekly reference data. Ansaldi et al. compared the ED data with the sentinel GP data following a daily approach and reached a similar specificity and sensitivity using a 5-day moving average algorithm. In general, it is difficult to compare validity between different SyS systems as factors, such as the chosen statistics/algorithms, represented population and outbreak characteristics, are influencing the performance.

In one of the seasons the ED data-based SyS system indicated the beginning of the influenza season 1 week earlier.
than the sentinel system. In the other season both data sources showed elevated influenza activity in the same week. This is in line with the results of other SyS systems based on the ED data that indicated a similar or earlier onset of 1–3 weeks for influenza compared with traditional surveillance data sources depending on the reference data sources used and statistical measures applied. The application of the same case definition in the two systems in Cantabria makes it possible to more precisely determine the early detection effect of the ED data compared with the sentinel GP data. This could be a reflection of the treatment-seeking behaviour in Spain that is characterized by a high number of patients utilizing EDs with conditions that would more appropriately be treated in a primary care setting.

For the season 2010–11, the health system organization and treatment-seeking behaviour of the population is also the likely explanation for the earlier wave seen in the ED. The period of high influenza activity began during the Spanish Christmas holidays (26 December–10 January) when most GP practices were closed and patients visited the ED instead. This highlights the usefulness of an ED-based surveillance system that is continuously operating during the entire year while the sentinel system based on GP practices is not. The beginning of the influenza season in season 2010–11 could only be detected in near real time by the ED SyS system. This suggests that ED data-based SyS can constitute an important supplement to the surveillance system to get a more complete picture of the influenza disease burden on the population. However, for general seasonal influenza surveillance, it needs to be taken into account that the populations visiting GPs and the ED are usually reflecting different populations, i.e. the general population, and more severe and acute, out-of-hours cases, respectively.

The influenza season 2010–11 in Santander also highlights the pressure of a high number of influenza patients on the resources of the ED, especially during holiday seasons when other health-care providers are closed. The implementation of a daily automated SyS system with locally adjusted algorithms and alert thresholds supports the decision-makers in the hospital to timely adjust their resources to the actual demand.

Once implemented, an automated SyS system can be a stable and flexible tool for timely monitoring of various public health events. Such a generic surveillance system can rapidly be applied during unexpected crisis situations or out-of-season epidemics such as the influenza pandemic in 2009. The Department of Health of the Autonomous Region of Cantabria officially endorsed the integration of ED data-based SyS as implemented in Santander as a supplementary surveillance information source in the region.

Limitations of this study
The time period of two seasons of influenza is quite short for a representative assessment. Furthermore, the analysis has been conducted retrospectively. A prospective analysis also of the experience of the regional public health authority during influenza seasons can bring further insight in the future.

The objectivity of physicians in diagnosing ILI could be influenced by the higher expectation of influenza case occurrence during the influenza season as suggested by Moore et al. We anticipate the multiple aspect case definition used in Santander to reduce false diagnoses to a minimum.

We applied one aberration detection algorithm that is used in other syndromic influenza surveillance systems and that is easy to implement but there are others that are equally often used such as regression analysis or moving averages. A statistical modelling of time series data with seasonal autoregressive moving average (SARIMA) algorithms could be suitable to our context as seasonal patterns and trends can be accounted for in a more precise manner.

The spatial distribution and spread of influenza can yield additional information for decision-making during an epidemic. Other studies have shown that syndromic data can be scanned for space-time clusters. These studies focused on large areas such as New York City, England or the Netherlands consisting of a large number of small-area postal code regions as basis for the analysis. Santander covers a comparatively small area with 12 postal codes for which we considered a spatial analysis not to provide an added value. This assumption could change in the future if the SyS system is enlarged to cover neighbouring areas in addition.

Conclusion
This case study shows the added value of SyS based on routine ED data to effectively complement sentinel surveillance for seasonal influenza at the local level in Spain. We recommend the local adjustment of the aberration detection algorithms and the definition of alert criteria per syndrome and data source to improve the performance of a SyS system.

Acknowledgement
The authors would like to thank the Instituto de Salud Carlos III for providing the reference data from sentinel GPs for the Autonomous Region of Cantabria.

Funding
This work was supported by the European Union, in the framework of the Public Health Programme (grant number 2007/208).
References


