In the last decade, syndromic surveillance has increasingly been used worldwide for detecting increases or outbreaks of infectious diseases that might be missed by surveillance based on laboratory diagnoses and notifications by clinicians alone. There is, however, an ongoing debate about the feasibility of syndromic surveillance and its potential added value. Here we present our perspective on syndromic surveillance, based on the results of a retrospective analysis of syndromic data from six Dutch healthcare registries, covering 1999–2009 or part of this period. These registries had been designed for other purposes, but were evaluated for their potential use in signal-ing infectious disease dynamics and outbreaks. Our results show that syndromic surveillance clearly has added value in revealing the blind spots of traditional surveillance, in particular by detecting unusual, local outbreaks independently of diagnoses of specific pathogens, and by monitoring disease burden and virulence shifts of common pathogens. Therefore we recommend the use of syndromic surveillance for these applications.

Background

In the last decade, syndromic surveillance has increasingly been implemented to detect and monitor infectious disease outbreaks, as early detection and control may well mitigate the impact of epidemics [1-3]. In the United Kingdom, for example, a telephone health helpline (NHS Direct) is used for syndromic surveillance [4]; in France, a syndromic surveillance system based on hospital emergency data has been deployed [4]; and in North America several syndromic surveillance systems exist using data such as telephone helpline calls [5] and hospital emergency department visits [2,6]. Traditional outbreak detection based on astute clinicians and laboratory diagnoses can have blind spots for emerging diseases, because patients reporting with common symptoms (e.g. pneumonia) associated with the disease may not alarm clinicians, and uncommon or new pathogens can remain undetected by laboratories (such as initially happened in the outbreak of severe acute respiratory syndrome (SARS) in 2003). Syndromic surveillance may reveal such blind spots of traditional surveillance by monitoring elevations of common symptoms or clinical diagnoses such as shortness of breath or pneumonia.

The increasing use of syndromic surveillance seems driven by two factors: (i) high-profile disease events (e.g. the 2001 anthrax attacks, 2003 SARS outbreak, the threat of a new influenza pandemic, excess mortality due to heat waves) stressing the need for improved early warning surveillance; and (ii) the increased availability of electronic healthcare data, making large-scale monitoring of non-specific health indicators increasingly feasible.

There is, however, an ongoing debate about the added value of syndromic surveillance. Some scepticism exists about the potential workload it may generate if used for real-time outbreak detection (i.e. if the system creates many false-positive signals) [7]. In the Netherlands, this debate has led to a research project to evaluate the potential value of syndromic surveillance for infectious disease surveillance and control, and to make recommendations for its implementation. The questions addressed were: (i) what syndromic data types track known dynamics of infectious diseases in the general population, and thus will also be likely to reflect emerging pathogen activity? (ii) can syndromic surveillance improve the monitoring of disease burden and/or detect shifts in the virulence of common pathogens? (iii) can syndromic surveillance detect local outbreaks that have a limited number of signals in time, independent of laboratory detection of the causative pathogens?

We addressed these questions by retrospectively analysing syndromic data from six Dutch healthcare registries, and also by ad hoc use of syndromic surveillance for upcoming infectious disease problems. To select
### Table 1
Syndromic data registries included in the syndromic surveillance evaluation study, the Netherlands

<table>
<thead>
<tr>
<th>Data type</th>
<th>Registry</th>
<th>Period evaluated</th>
<th>Population coverage (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Syndrome information&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Data analysed</th>
<th>International coding system</th>
<th>Prospective implementation of surveillance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absenteeism</td>
<td>National absenteeism registry by Statistics Netherlands (CBS) [8]</td>
<td>2002–03</td>
<td>80 (of the working population of 8 million)</td>
<td>Employees reported sick, no further medical information</td>
<td>Sick leave reports by employers</td>
<td>–</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>General practitioner consultations</td>
<td>Netherlands Information Network of General Practice (LINH, by NIVEL-the Netherlands Institute for Health Services Research) [9]</td>
<td>2001–04</td>
<td>1–2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Symptoms and diagnoses indicating infectious disease</td>
<td>Symptoms and diagnoses recorded in general practice or telephone consultations, and home visits</td>
<td>International Classification of Primary Care (ICPC)</td>
<td>Real-time system currently being implemented</td>
</tr>
<tr>
<td>Pharmacy prescriptions</td>
<td>Foundation for Pharmaceutical Statistics (SFK) [10]</td>
<td>2001–03</td>
<td>85</td>
<td>Prescribed medications indicating infectious disease</td>
<td>Prescription medications dispensed in Dutch pharmacies</td>
<td>Anatomical Therapeutic Chemical Classification System (ATC)</td>
<td>Currently monthly data updates are feasible in ad hoc public health situations</td>
</tr>
<tr>
<td>Hospitalisations</td>
<td>Dutch National Medical Register (LMR) by Dutch Hospital Data (DHD) [11]</td>
<td>1999–2007</td>
<td>99</td>
<td>General symptoms and diagnoses and specific biological agent diagnoses</td>
<td>Discharge and secondary diagnoses and date of hospitalisation</td>
<td>International Classification of Diseases, Ninth, Revision Clinical Modification (ICD-9-CM)</td>
<td>No prospective implementation possible in the short term (annual data updates will continue)</td>
</tr>
<tr>
<td>Laboratory submissions (negative and positive results)</td>
<td>National Infectious Diseases Information System (ISIS-MML) [12]</td>
<td>2001–04</td>
<td>16</td>
<td>Submissions for specific microbiological diagnostic tests</td>
<td>Laboratory submission requests for diagnostic testing</td>
<td>–</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

NA: not applicable.

<sup>a</sup> Calculated as a percentage of the total population (16.3 million in 2006 [14]), unless otherwise indicated.

<sup>b</sup> Detailed syndrome definitions available from the authors on request.

<sup>c</sup> The laboratory submissions registry (ISIS-MML) and the national absenteeism registry ceased to exist during our study.

<sup>d</sup> The GP registry coverage will increase to 5% in the next few years as part of the Surveillance Network Netherlands (SUNN), which is predominantly focused on influenza surveillance [15].
### Table 2

Tracking of infectious disease dynamics using three syndromes and six data types from healthcare registries, syndromic surveillance evaluation study, the Netherlands

<table>
<thead>
<tr>
<th>Data types</th>
<th>Respiratory syndromes</th>
<th>Gastroenteritis syndromes</th>
<th>Neurological syndromes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absenteeism</td>
<td>Winter peaks concurrent with peaks in influenza virus, RSV and other respiratory pathogens; 68% of variations explained by respiratory pathogens; 2 weeks ahead of RSV, 4–5 weeks ahead of influenza [19]</td>
<td>Not evaluated&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Not evaluated&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>General practitioner consultations</td>
<td>Winter peaks concurrent with peaks in influenza, RSV and other respiratory pathogens; 86% of variations explained by respiratory pathogens, 1 week behind RSV, 1–2 weeks ahead of influenza [19]</td>
<td>Winter peaks and summer peaks concurrent with rotavirus and <em>Shigella/Salmonella/Campylobacter</em> peaks, respectively; 29% of variations explained by gastroenteral pathogens (51% for those aged 0–4 years, two weeks ahead of rotavirus) [20]; an increase in winter of 2002/03 possibly related to norovirus activity [21]</td>
<td>No clear reflection of known disease dynamics</td>
</tr>
<tr>
<td>Pharmacy prescriptions</td>
<td>Winter peaks concurrent with peaks in influenza, RSV and other respiratory pathogens; 80% of variations explained by respiratory pathogens; 1 week behind RSV, 0–2 weeks ahead of influenza [19]</td>
<td>Relatively low winter peaks and higher summer peaks concurrent with rotavirus and <em>Shigella/Salmonella/Campylobacter</em> peaks, respectively</td>
<td>Not evaluated&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hospitalisations</td>
<td>Winter peaks concurrent with peaks in influenza, RSV and other respiratory pathogens; 84% of variations explained by respiratory pathogens; in concurrence with RSV, 1–2 weeks ahead of influenza [19]</td>
<td>Relatively high winter peaks and lower summer peaks concurrent with rotavirus and <em>Shigella/Salmonella/Campylobacter</em> peaks, respectively</td>
<td>The general neurological syndrome did not clearly reflect known disease dynamics. A viral neurological syndrome showed summer peaks concurrent with enterovirus peaks: 62% of its variations was explained by enterovirus notifications [22]</td>
</tr>
<tr>
<td>Laboratory submissions (negative and positive results)</td>
<td>Winter peaks concurrent with peaks in influenza, RSV and other respiratory pathogens; 61% of variations explained by respiratory pathogens; 2 weeks behind RSV, 0–1 week ahead of influenza [19]</td>
<td>Relatively low winter peaks and higher summer peaks concurrent with rotavirus and <em>Shigella/Salmonella/Campylobacter</em> peaks, respectively</td>
<td>No clear reflection of known disease dynamics</td>
</tr>
<tr>
<td>Mortality</td>
<td>Winter peaks concurrent with peaks in influenza, RSV and other respiratory pathogens; 78% of variations explained by respiratory pathogens; 3 weeks behind RSV, 0–1 week ahead of influenza [19]</td>
<td>No obvious reflection of known seasonal pathogen activity; an increase in winter 2002/03 possibly related to norovirus activity [21]</td>
<td>No clear reflection of known disease dynamics</td>
</tr>
</tbody>
</table>

RSV: respiratory syncytial virus.

The table summarises per data and syndrome type whether syndrome peaks concurred with pathogen peaks, what percentage of the syndrome variations is explained by variations in pathogen counts, and what the differences in timeliness were between the syndrome and pathogen data. The latter are assessed by the optimised lags of pathogen counts in time-series models that explain the syndrome variation [19].

<sup>a</sup> The absenteeism data lacked medical information, but its time series reflected respiratory pathogen activity; therefore the other syndromes were not evaluated for this registry.

<sup>b</sup> No data on pharmacy prescriptions specific for neurological conditions were available for analysis.
potential syndromic data sources, we asked Dutch healthcare registry owners to provide information on predefined criteria (coverage, timeliness of data entry and potential for transition to real-time data availability). Table 1 shows the registries included in the study, with data on work absenteeism, general practitioner (GP) consultations, pharmacy prescriptions, laboratory submissions, hospitalisations and mortality. Data were available for 1999–2009 or part of this period.

On the basis of available literature cited in PubMed on bioterrorism and natural infectious disease threats, we selected syndromes that were expected to reflect the clinical presentations of both high-threat (i.e. capable of causing major outbreaks of severe illness) and common pathogens [16,17]. This not only makes it possible to use common pathogen activity as a test case for these syndromes, but also implies that emergence of the high-threat pathogens concerned will be relatively difficult to recognise by clinicians. We selected respiratory syndromes (e.g. for high-threat pathogens such as Bacillus anthracis or a new pandemic influenza variant), gastroenteritis syndromes (e.g. caused by Vibrio cholerae infection) and neurological syndromes (e.g. caused by West Nile virus infection). The syndromes were defined for each registry, guided by a list of syndrome definitions defined by the United States Centers for Disease Control and Prevention [18] and experts in infectious diseases and medical microbiology at the Rijksinstituut voor Volksgezondheid en Milieu (National Institute for Public Health and the Environment, RIVM). The syndromes were then evaluated per registry for their potential use in signalling infectious disease dynamics and outbreaks.

In this article, we present our perspective on the added value of syndromic surveillance for infectious disease surveillance and control, based on the results of our evaluation study and in light of the literature up to and including 2010.

Main findings of syndromic surveillance evaluation
Tracking infectious disease dynamics in the general population
The first question we addressed was to what extent trends in respiratory, gastroenteritis and neurological syndromes in the various registries reflect known pathogen activity, as measured by counts of detected pathogens (available from routine laboratory surveillance). This indicates whether these registries have the potential to reflect emerging pathogen activity (Table 2).

Of the three syndromes, respiratory syndromes were most closely associated with laboratory pathogen counts (Table 2), displaying higher levels in winter, which corresponded to higher counts of respiratory pathogens [19]. Up to 86% of the weekly syndrome variations (i.e. variance) in time were explained by weekly variations in respiratory pathogen counts, particularly of influenza viruses and respiratory syncytial virus (RSV), which is in line with other studies [23,24]. However, the respiratory syndromes in our study were zero to five weeks ahead of laboratory counts of influenza viruses, suggesting better timeliness of these syndromes. For RSV, the pathogen counts were concurrent with respiratory syndromes from hospitalisation registry data, which would be expected as most RSV tests are performed on hospitalised young children [25,26]. Most respiratory syndromes from other registry data lagged behind the RSV counts, which suggests that young children are affected relatively early in the annual RSV season.

The gastroenteritis syndromes showed winter peaks concurrent with increased rotavirus activity, and summer peaks concurrent with peaks in Shigella, Campylobacter and Salmonella activity (Table 2). Variation in the reporting of gastroenteritis syndromes explained by pathogen counts was lower (29–40%) than in the respiratory syndromes, although it increased up to 85% when limiting the analysis to young children, with the syndromes’ counts one to two weeks ahead of the laboratory rotavirus counts [20].

The reported general neurological syndromes did not clearly reflect known patterns of pathogen activity (Table 2). However, a more specific viral neurological syndrome – unexplained viral meningitis syndrome – in the hospitalisation registry data did: 62% of the variation in the reporting of this syndrome was explained by known seasonal enterovirus activity, suggesting that elevated levels of unexplained viral meningitis indicate undiagnosed enterovirus infections [22].

The general practitioner consultations, pharmacy prescriptions, hospitalisations and mortality registry data thus showed good performance in timely tracking of respiratory and/or gastrointestinal disease, and the hospital registry data also showed moderate performance for neurological disease (Table 2). The advantage of using these four complementary registries together would be that they cover mild to very severe morbidity. The absenteeism registry data seemed most timely (ahead of laboratory surveillance data), but showed only moderate performance in tracking respiratory disease. This could be due to the fact that medical information is not available in this registry, and thus the data are a mix of all kinds of disease, although respiratory disease is clearly reflected in the time-series pattern. The laboratory submissions registry data showed, at most, a moderate performance for the three syndromes evaluated.

Monitoring disease burden and detecting virulence shifts
The second research question we addressed was whether syndromic surveillance improves the monitoring of disease burden and detects shifts in the virulence of common pathogens. We evaluated this by relating time series of syndromic surveillance data with
pathogen-specific surveillance data to quantify the disease burden due to common pathogens over time. We found a clear association over time of norovirus laboratory surveillance data with mild-to-severe morbidity and even deaths in elderly people, observed in recent years, coinciding with emergence of new norovirus variants [21]. The emergence of these variants had been suspected but could not be assessed by any other routine surveillance system. In addition, for influenza we detected previously unknown shifts in the annual numbers of hospitalisations and deaths related to the number of influenza-like illness (ILI) cases, coinciding with shifts in the antigenicity of circulating viruses [27]. Such analyses can also be used for investigating the severity of pandemic influenza A(H1N1)2009 infection compared with that of seasonal influenza [28].

Detecting local outbreaks
The third question we addressed was whether syndromic surveillance can detect unexpected disease outbreaks in a timely manner. For this purpose, analysis of aggregated nationwide data may not be very sensitive: the large volume of the data (e.g. tens of thousands of respiratory syndrome hospitalisations per year) makes it impossible to detect outbreaks when they are still small. Local detection of syndrome elevations using a space–time algorithm might signal emerging outbreaks much sooner [29]. To test this, we used known outbreaks of Legionnaires’ disease as positive controls of realistic severe respiratory disease outbreaks due to uncommon or new pathogens that may not be detected by traditional surveillance in a timely manner. Simulating prospective surveillance, we were able to timely detect these known outbreaks in syndromic hospital data using space–time scan statistics [29]. The fact that the overall alarm rate was modest (a mean of five local clusters detected per year) suggests that syndromic surveillance of hospitalisation data for respiratory disease can indeed be a useful early-warning tool for local outbreak detection. Using the same approach, previously unknown disease clusters plausibly due to Q fever were detected [30], thus illustrating that on some occasions syndromic surveillance can identify outbreaks that otherwise would remain undetected. These analyses were motivated by the clinical detection of a large Q fever outbreak in 2007 and the subsequent years, which raised the question whether smaller outbreaks might have preceded the 2007 outbreak. Real-time detection and investigation of these previously unknown clusters could possibly have led to earlier awareness of increased Q fever activity.

Assessing the absence or limited size of unusual disease events
In public health practice, besides timely detection of unusual outbreaks, being able to assess and communicate the absence or limited size of unusual disease events can also be important. For example, Blendon et al. suggested that better communication to the public during the 2003 SARS outbreak might have prevented economic loss due to unnecessary precautions taken by the public (e.g. many people stayed away from crowded places, even in areas with a relatively low level of spread of the virus) [31]. Also during high-profile public events (e.g. the Olympics or G8 summits) [32–33], syndromic surveillance will mainly confirm the absence of major, unusual disease outbreaks, since such outbreaks are rare events.

We also examined the value of syndromic data in assessing the absence or limited size of unusual disease triggered by specific public health concerns. For West Nile virus (WNV) infection, enhanced surveillance was established in the Netherlands by laboratory testing of cerebrospinal fluid (CSF) from patients with unexplained viral meningitis/encephalitis [22]. None of the CSFs collected in 2002 to 2004 tested positive for WNV, but the probability that WNV was indeed absent in the country could only be assessed from the annual count of unexplained viral meningitis/encephalitis cases (as a denominator in relation to the number of CSF samples tested). For hepatitis E and Ljungan virus infections, we inspected time series of unexplained hepatitis and abortion/perinatal death, respectively, and found no signs of emerging activity of these viruses. Rumours about a continuing increase of impetigo in children were countered by inspection of a time series of GP consultations for the infection.

Other spin-offs of syndromic surveillance
In addition to the above described applications, other uses of syndromic surveillance were illustrated during the 2009 influenza A(H1N1) pandemic. We used respiratory syndromic data on hospitalisations and GP consultations to plan the diagnostic capacity that would be needed if a larger proportion of the persons with respiratory symptoms would be tested – as is the case in the early stages of a pandemic [34]. Also early in the pandemic, the reaction of the public to media reports on pandemic influenza was illustrated by sharp elevations in the number of oseltamivir prescriptions [35]. This information was used to urge physicians to exercise restraint in prescribing oseltamivir, in order to decrease the risk of oseltamivir shortage and viral resistance later in the pandemic.

Data requirements
The results of our project suggest specific data requirements for successful syndromic surveillance. Data quality is important for all applications of syndromic surveillance, but probably mostly for local outbreak detection. Here, relatively small artefacts – for example, duplicate details of the same patient in one registry – can result in false alarms, as we experienced when using hospital data for space–time cluster detection [29,30]. In a real-time setting (e.g. daily or weekly data updates), reporting delays can also lead to data artefacts and false alarms, if, for example, there is a delay in hospitals submitting their data [36]. In addition to having few data artefacts, data need to be representative, and for local outbreak detection, they also need to have a high coverage (preferably close
to 100%) to be able to timely detect local outbreaks in any region. By using data with relatively low coverage levels, sensitivity for local outbreaks obviously will be reduced [37,38]. Nordin et al. used simulated anthrax attack data, and integrated the simulated data into actual physicians’ visit data to show that the sensitivity for detecting respiratory outbreaks resulting from bioterrorism was not very high [37]. However, the authors evaluated a maximum system coverage of only 36% of the population. In another study, Balter et al. reported that a syndromic surveillance system in New York City sometimes missed several gastroenteritis outbreaks due to data quality (e.g. miscoding of patients’ chief complaints) and coverage problems (e.g. some hospitals did not participate in the system) [38].

For effective signal verification, sufficient information on individual patients’ characteristics and concurrent laboratory trends has to be available to identify possible causes of generated signals. For example, we interpreted local respiratory syndrome clusters in relation to local influenza or RSV activity: if the age distribution of cases reflected the usual pattern for these viruses, we regarded further investigation unnecessary [29]. Also, the rise in oseltamivir prescriptions early in the 2009 influenza A(H1N1) pandemic could be ascribed to the ‘worried well’, because laboratory surveillance showed that influenza virus activity had not increased [35]. Without such verification options, the value of syndromic surveillance is limited [38].

Cost-effectiveness of real-time surveillance systems

An important question is whether syndromic surveillance is cost effective. Events such as a bioterrorist attack, a SARS epidemic or an influenza pandemic are rare and the question arises how much of the public health budget should be spent on a detection system for such rare events.

The costs of a surveillance system can be easily estimated. Studies that report the operating costs associated with real-time syndromic surveillance found annual operating costs ranging from US$ 130,000–150,000 to US$ 280,000 [39]. However, estimating its benefits is less obvious. Kaufmann et al. reported that the economic damage caused by a bioterrorist attack can amount to millions or even billions of dollars [40]. The SARS epidemic in 2003 and the influenza pandemic in 2009 showed that the economic damage caused by naturally occurring outbreaks can be similarly high [41,42]. If similar disease events emerge every few years, and syndromic surveillance leads to earlier detection and control of such outbreaks, then the benefits of syndromic surveillance are likely to outweigh its costs. The question here is whether earlier detection would indeed lead to control or at least reduced impact of a new disease, for instance, SARS or influenza A(H1N1)2009 infection. Simulation studies could help to further evaluate which specific types of major disease events syndromic surveillance could probably lead to interventions that limit the economic damage.

Possibly just as important as the benefits arising from earlier detection and control is the downscaling of unnecessary interventions during ongoing outbreaks. This requires quick assessment of the limited size and severity of outbreaks. For example, if the severity of a new pandemic can be quickly assessed – as the World Health Organization (WHO) requires [43] – by reliable syndromic hospital surveillance of severe respiratory infections, costly interventions such as quarantine and prophylactic treatment or vaccination could be down-scaled or stopped earlier if the disease is only mild.

In the Netherlands, prospective surveillance has now started for crude mortality data, with weekly data collection and analysis since the 2009 influenza pandemic. The existing mortality registry allows prospective implementation at relatively low extra cost. Real-time data collection is currently also being implemented for the Dutch GP registry (Table 1). Including hospital data and other data types in future syndromic surveillance systems may also be feasible at limited cost, if the data collection can be integrated into already planned real-time, future data infrastructures such as the Dutch national health-information-exchange system [44].

Recommendations

On the basis of our evaluation, we recommend the use of syndromic surveillance to reveal blind spots of traditional surveillance, in particular by detecting unusual, local outbreaks independently of laboratory diagnoses of specific pathogens, and by monitoring disease burden and virulence shifts of common pathogens.

Our results are mostly based on retrospective analysis of syndromic data of high quality and coverage. If prospective collection of such syndromic data is not feasible, real-time early warning for local outbreaks should not be performed, since true outbreaks will probably be missed while at the same time numerous false alarms will be generated. For real-time early warning, sufficient laboratory and epidemiological information is needed, in order to be able to quickly verify possible causes of syndromic signals, and thus recognise relevant signals that might need a response. Retrospective analyses as performed in our evaluation can validate the relevant data and analyses before prospective implementation of a syndromic surveillance system.

Acknowledgements

We thank Statistics Netherlands (CBS), the Dutch Foundation for Pharmaceutical Statistics (SFK), ISIS-labs, Dutch Hospital Data and Prismant (LMR), Netherlands Institute for Health Services Research (NIVEL) (regarding both the Netherlands Information Network of General Practice (LINH) and the Continuous Morbidity Registration (CMR) sentinel stations) for providing data, and the members of the Dutch Working Group on Clinical Virology, for providing access to the registry on weekly positive diagnostic results.
