Validation of Syndromic Surveillance for Respiratory Pathogen Activity

Cees van den Wijngaard,* Liselotte van Asten,* Wilfrid van Pelt,* Nico J.D. Nagelkerke,† Robert Verheij,‡ Albert J. de Neeling,* Arnold Dekkers,* Marianne A.B. van der Sande,* Hans van Vliet,* and Marion P.G. Koopmans*

Syndromic surveillance is increasingly used to signal unusual illness events. To validate data-source selection, we retrospectively investigated the extent to which 6 respiratory syndromes (based on different medical registries) reflected respiratory pathogen activity. These syndromes showed higher levels in winter, which corresponded with higher laboratory counts of *Streptococcus pneumoniae*, respiratory syncytial virus, and influenza virus. Multiple linear regression models indicated that most syndrome variations (up to 86%) can be explained by counts of respiratory pathogens. Absenteeism and pharmacy syndromes might reflect nonrespiratory conditions as well. We also observed systematic syndrome elevations in the fall, which were unexplained by pathogen counts but likely reflected rhinovirus activity. Earliest syndrome elevations were observed in absenteeism data, followed by hospital data (+1 week), pharmacy/general practitioner consultations (+2 weeks), and deaths/laboratory submissions (test requests) (+3 weeks). We conclude that these syndromes can be used for respiratory syndromic surveillance, since they reflect patterns in respiratory pathogen activity.

Early warning surveillance for emerging infectious disease has become a priority in public health policy since the anthrax attacks in 2001, the epidemic of severe acute respiratory syndrome in 2003, and the renewed attention on possible influenza pandemics. As a result, new surveillance systems for earlier detection of emerging infectious diseases have been implemented. These systems, often labeled "syndromic surveillance," benefit from the increasing timeliness, scope, and diversity of health-related registries (*1–6*). Such alternative surveillance uses symptoms or clinical diagnoses such as "shortness of breath" or "pneumonia" as early indicators for infectious disease. This approach not only allows clinical syndromes to be monitored before laboratory diagnoses, but also allows disease to be detected for which no additional diagnostics were requested or available (including activity of emerging pathogens). Our study assessed the suitability of different types of healthcare data for syndromic surveillance of respiratory disease.

We assumed that syndrome data—to be suitable for early detection of an emerging respiratory disease—should reflect patterns in common respiratory infectious diseases (*7–10*). Therefore, we investigated the extent to which time-series of respiratory pathogens (counts per week in existing laboratory registries) were reflected in respiratory syndrome time-series as recorded in 6 medical registries in the Netherlands. We also investigated syndrome variations that could not be explained by pathogen counts. As an indication for syndrome timeliness, we investigated the delays between the syndrome and pathogen time-series.

Methods

Syndrome Data Collection and Case Definitions

We defined syndrome data as data in health-related registries that reflect infectious disease activity without identifying causative pathogen(s) or focusing on pathogen-specific symptoms (such as routine surveillance data for influenza-like illness [*11*] or surveillance of acute flaccid paralysis for polio [*12*]).

Registries for syndrome data were included if they met the following criteria: 1) registration on a daily basis; 2) availability of postal code, age, and sex; 3) availability

*National Institute for Public Health and the Environment, Bilthoven, the Netherlands; †United Arab Emirates University, Al-Ain, United Arab Emirates; and ‡Netherlands Institute of Health Services Research, Utrecht, the Netherlands

of retrospective data (≥2 years); and 4) (potential) real-time data availability.

Six registries were selected (Table 1) that collected data on work absenteeism, general practice (GP) consultations, prescription medications dispensed by pharmacies, diagnostic test requests (laboratory submissions) (*13*), hospital diagnoses, and deaths. In all registries, data were available for all or a substantial part of 1999–2004. For the GP, hospital, and mortality registry, definition of a general respiratory syndrome was guided by the case definitions and codes found in the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM), as selected by the Centers for Disease Control and Prevention (Atlanta, GA, USA) (www.bt.cdc.gov/surveillance/syndromedef). For the laboratory submissions and the pharmacy syndrome, we selected all data that experts considered indicative of respiratory infectious disease (for detailed syndrome definitions, see online Technical Appendix, available from www.cdc.gov/EID/content/14/6/917-Techapp.pdf).

Respiratory Pathogen Counts

As a reference for the syndrome data, we included specific pathogen counts for 1999–2004 from the following sources: 1) Weekly Sentinel Surveillance System of the Dutch Working Group on Clinical Virology (which covers 38%–73% of the population of the Netherlands [*14*] for routine laboratory surveillance of respiratory syncytial virus [RSV], influenza A virus, influenza B virus, rhinovirus, *Mycoplasma pneumoniae*, parainfluenza virus, enterovirus, and adenovirus); 2) 6 regional public health laboratories for

Table 1. Registries from which syndrome data were obtained, the Netherlands, 1999–2004*

Data type	Period	% Coverage†	Respiratory syndrome definitions‡	Analyzed data	International code system	Registry
Absenteeism	2002–2003	80§	Reported sick employees; no further medical information	Sick leave reports of employees	–	Statistics Netherlands (CBS), www.cbs.nl
General practice consultations	2001–2004	1–2	Symptoms and diagnoses indicating respiratory infectious disease	Symptoms and diagnoses recorded in practice or telephone consultations and in home visits	ICPC	Netherlands Information Network of General Practice (LINH), www.nivel.nl/linh
Pharmacy dispensations	2001–2003	85	Prescribed medications indicative for respiratory infectious disease	Prescription medications dispensed in Dutch pharmacies, coded according to the WHO ATC classification	ATC	Foundation for Pharmaceutical Statistics, http://www.sfk.nl
Hospitalization	1999–2004	99	General respiratory symptoms/diagnoses; specific respiratory biologic agent diagnoses	Discharge and secondary diagnoses, date of hospitalization	ICD-9-CM	Dutch National Medical Register (LMR)
Laboratory submissions¶	2001–2004 (1999–2000 excluded due to unstable coverage)	16	All submissions for microbiologic diagnostic tests on respiratory materials; all submissions for serologic testing on known specific respiratory pathogens; all submissions for *Legionella* or *Streptococcus pneumoniae* antigen tests on urine	Laboratory submission requests for diagnostic testing	–	National Infectious Diseases Information System (ISIS) (*13*)
Mortality	1999–2004	100	General respiratory symptoms/diagnoses; specific respiratory biologic agent diagnoses	Date of death, primary cause of death, complicating factors, other additional causes of death	ICD-10	CBS

*ICPC, International Classification of Primary Care; WHO, World Health Organization; ATC, Anatomic Therapeutic Chemical Classification System; ICD-9-CM, International Classification of Diseases, 9th revision, Clinical Modification; ICD-10, International Classification of Diseases, 10th revision.
†Percentage of total population, 16.3 million.
‡For detailed syndrome definitions and codes, see online Technical Appendix, available from www.cdc.gov/EID/content/14/6/917-Techapp.pdf.
§Percentage of working population, 8 million.
¶Diagnostic test requests with both negative and positive results.

respiratory disease–related counts of *Streptococcus pneumoniae* (data in 2003–2004 were interpolated for 2 laboratories during short periods of missing data; total coverage 24%); and 3) national mandatory notifications of pertussis. The networks for respiratory pathogen counts are other networks than the earlier described laboratory submissions network for syndrome data.

Data Analysis and Descriptive Statistics

Data were aggregated by week and analyzed by using SAS version 9.1 (SAS Institute Inc., Cary, NC, USA). For the GP, pharmacy, and laboratory submissions registries, we expressed the respiratory counts as a percentage of total weekly counts to adjust for the influence of holidays and, for laboratory submissions, changes in the number of included laboratories over time. By looking at the graphs, we explored the relationship between the time-series of respiratory pathogens and syndromes and calculated Pearson correlation coefficients.

Linear Regression Models

To investigate whether the respiratory syndromes reflect patterns in respiratory pathogen counts, we constructed multiple linear regression models. These models estimated respiratory syndrome levels at a certain time with, as explanatory variables, the lagged (range of –5 to +5 weeks) pathogen counts as explanatory variables. We used linear regression of the untransformed syndrome to estimate the additive contributions of individual pathogens to the total estimated syndrome. We assumed a constant syndrome level attributable to factors other than the respiratory pathogens and constant scaling factors for each of the lagged pathogens. A forward stepwise regression approach was used, each step selecting the lagged pathogen that contributed most to Akaike's information criterion of model fit (*15*). Each pathogen entered the model only once and only if it contributed significantly ($p<0.05$). Negative associations (e.g., between enteroviruses, which peak in summer, vs. respiratory syndromes, which peak in winter) were excluded to avoid noncausal effects.

To discriminate between primary and secondary infections by *S. pneumoniae* (as a complication of respiratory virus infection) (*16–19*), we used the residuals from regressing *S. pneumoniae* counts on other pathogens as the variable for *S. pneumoniae* (instead of its counts) for all the earlier described models for respiratory syndromes.

We checked for autocorrelation in the residuals of the models with hierarchical time-series models (using SPLUS 6.2) (*20*,*21*). We calculated R^2 values to estimate to what extent respiratory pathogen counts explain variations in syndromes. To explore to what extent seasonal variation could be a confounder, we also calculated R^2 values of the models after adding seasonal variables (sine and cosine terms) and R^2 values for seasonal terms alone. We also investigated the pathogen-specific effects in the models, by calculating the standardized parameter estimates before and after adding seasonal terms.

The models were used to estimate the expected syndrome level with 95% upper confidence limits (UCLs). We considered distinct syndrome elevations that exceeded the UCLs, as unexplained by the models (for model details, see online Technical Appendix).

Timeliness

We investigated the timeliness of the registry syndromes in 2 ways: 1) as a measure of differences in timeliness between registries, we evaluated the time delays of the syndromes relative to each other by calculating for each of the syndromes the time lag that maximized Pearson correlation coefficient with the hospital registry (as a reference); 2) by estimating the time delays between each of the syndromes and the lagged pathogens included in its regression model.

Results

Data Exploration and Descriptive Statistics

Respiratory syndrome time series were plotted for all registries (Figure 1). The Christmas and New Year holidays coincided with peaks and dips in the pharmacy and absenteeism syndromes (not shown). Because these results were probably artifacts, we smoothed these yearly peaks and dips and censored them in the analyses performed on the absenteeism registry, in which they had a strong influence on outcomes. For all registries, the respiratory syndromes demonstrated higher levels of activity in winter, which overlapped or coincided roughly with the seasonal peaks of influenza A, influenza B, RSV, and (albeit less pronounced) *S. pneumoniae* laboratory counts (Figure 1). Infections with parainfluenza virus, *M. pneumoniae*, adenovirus, and rhinovirus were detected slightly more frequently during winter (data not shown). *Bordetella pertussis* and enterovirus showed seasonal peaks only in summer (data not shown).

The seasonal peaks in laboratory counts of influenza A, influenza B, and RSV corresponded with peaks in the GP, pharmacy, and hospital syndromes. Other syndromes did have less obvious correspondence. Each year, around October, the respiratory syndrome showed a peak in the GP (2001–2004), pharmacy (2001–2003), and absenteeism (2002–2003) registries (Figure 1, panels A–C) that was observed neither for the other registries nor in any of the laboratory pathogens.

We calculated Pearson correlation coefficients between the different unlagged time series of respiratory pathogens and syndromes (Table 2). Syndrome time series in all reg-

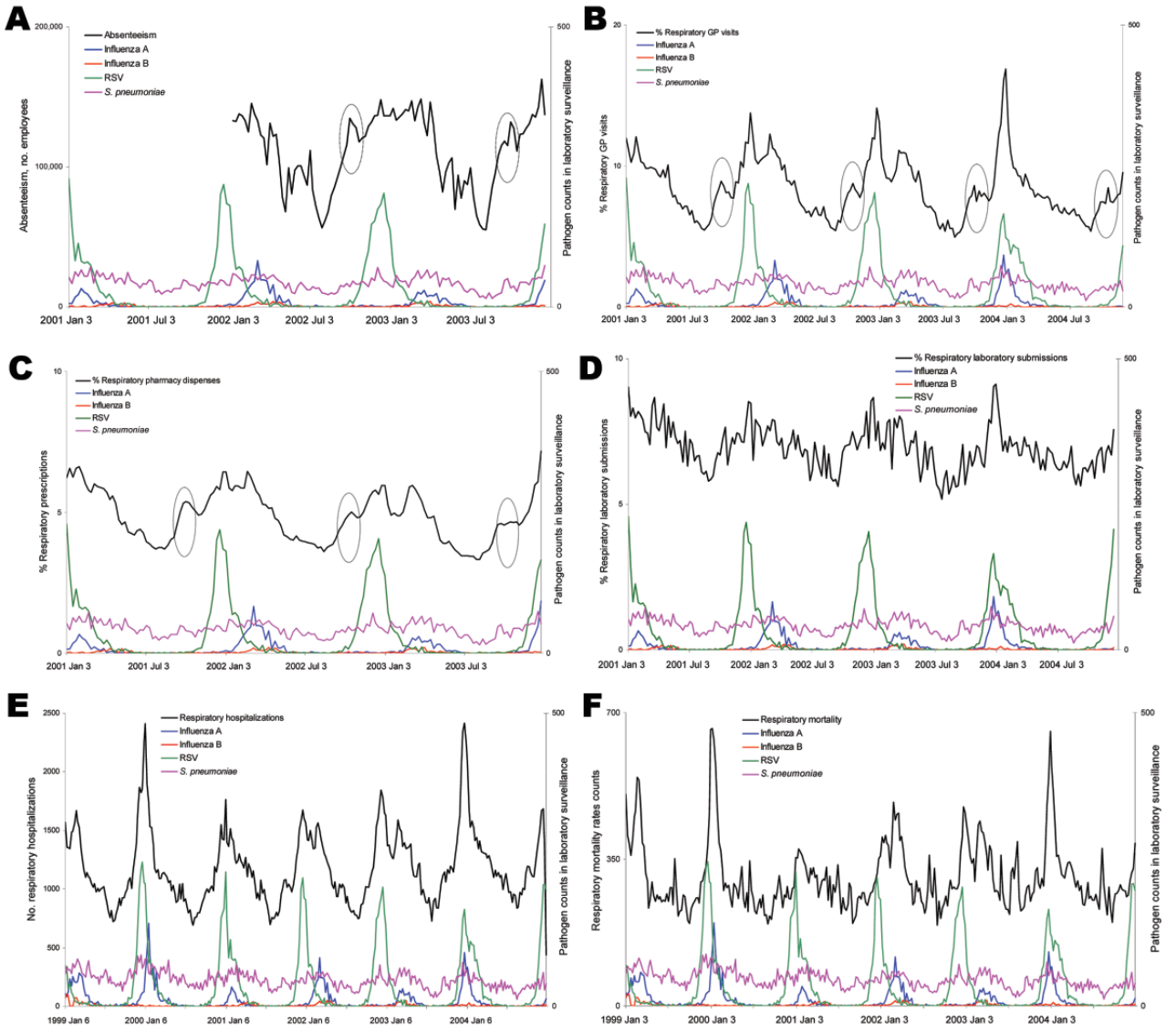

Figure 1. Respiratory syndrome time series and laboratory pathogen counts in the Netherlands. Respiratory syndromes were defined for the 6 registries defined in Table 1: A) absenteeism, B) general practice (GP) consultations, C) pharmacy, D) laboratory submissions, E) hospitalizations, and F) mortality counts. Pathogens plotted were respiratory syncytial virus (RSV), influenza A, influenza B, and *Streptococcus pneumoniae* [1999–2004 or part of this period, panels A–C]. Recurrent unexplained syndrome elevations in October are circled. Pathogen counts are daily counts of pathogens found in laboratory surveillence.

istries correlated strongly with *S. pneumoniae* (unadjusted total counts). The hospital, GP, pharmacy, and laboratory submissions data strongly correlated with RSV and influenza A counts (Table 2). Mortality data correlated strongly with influenza A (r = 0.65) and influenza B (r = 0.50) infections. The highest correlations between pathogen time series were between *S. pneumoniae* and the other pathogens (up to 0.51 with influenza A, Table 3).

Linear Regression Models

Table 4 presents, for each registry, the time lag (in weeks) that maximized the model fit of regressing syndrome on pathogens. For the GP, hospital, mortality, and pharmacy data, the respiratory pathogens explained the syndrome variation very well (78%–86%). Variations in the absenteeism syndrome could be explained for 68% by variations in the pathogen counts. Although the laboratory submissions syndrome had the lowest explained variance, still 61% of the variations in this syndrome were explained by variations in pathogen counts. Hierarchical time-series models did not show significant autocorrelation in the residuals of the models with pathogen counts as explanatory variables (*20*,*21*).

When seasonal terms were added to the model, the variations in the mortality syndrome were just as well ex-

Table 2. Pearson correlation coefficients between time series of syndromes and laboratory pathogen counts, the Netherlands, 1999–2004*†

Pathogen	Hospital	GP	Mortality	Pharmacy	Laboratory submissions	Absenteeism
RSV	**0.74**	**0.67**	0.41	**0.58**	**0.53**	0.47
Influenza A	**0.57**	**0.61**	**0.65**	**0.60**	0.47	0.35
Influenza B	0.31	0.39	**0.50**	0.42	0.34	0.33
Streptococcus pneumoniae	**0.73**	**0.71**	**0.56**	**0.75**	**0.58**	**0.69**
Rhinovirus	0.33	0.34	0.33	0.33	NS	0.35
Parainfluenza	0.20	NS	NS	NS	0.25	NS
Adenovirus	0.37	0.35	0.33	0.36	NS	0.34
Enterovirus	–0.65	–0.66	–0.59	–0.61	–0.57	–0.51
Mycoplasma pneumoniae	0.13	0.27	0.25	0.39	0.32	0.26
Bordetella pertussis	NS	NS	NS	NS	NS	NS

*GP, general practice; RSV, respiratory syncytial virus; NS, nonsignificant. Correlations ≥0.50 in **boldface**; p≥0.05.
†Unlagged.

plained as by the model with only pathogen counts (Table 5; R^2 remains 78%), while by the model with only seasonal terms, the explained variance was much lower (only 52%, Table 5). For the hospitalizations, laboratory submissions, and GP data, only slightly more syndrome variation was explained by adding seasonal terms. With only seasonal terms, the explained variance for these syndromes was clearly lower than with only pathogens in the models (8%–11% lower, Table 5). However, for the absenteeism and, to a lesser extent, the pharmacy data, the model with both pathogen and seasonal terms clearly explained more syndrome variations (Table 5, absenteeism 68% vs. 80%; pharmacy 80% vs. 87%). Furthermore, for the absenteeism data, the model with only seasonal terms had an even higher R^2 than the model with only pathogens, whereas for the pharmacy data, the R^2 with only seasonal terms was only slightly lower (3%, Table 5).

Table 6 shows that for mortality, hospitalizations, laboratory submissions, and GP data, the pathogens with the highest effect clearly were RSV, influenza A, and influenza B, with no or only modest decline in standardized parameter estimates after adding seasonal terms. For the GP and hospital data, some pathogens became insignificant after seasonal terms were added (GP: rhinovirus and adenovirus; hospital: parainfluenza virus). For the pharmacy data, half of all pathogen variables became insignificant after seasonal terms were added, whereas for the absenteeism data, almost all pathogens became insignificant (Table 6).

Several syndrome observations exceeded the 95% UCLs of the models (0–10/registry/year), which indicates that those syndrome observations deviated strongly from model predictions. The recurrent elevation in October of the absenteeism, GP, and pharmacy syndrome several times exceeded the UCLs (October 2001: pharmacy and GP; 2002: absenteeism; 2003: GP, absenteeism; not shown), which indicated that the model could not explain these elevations.

Timeliness

In Figure 2, for each registry, the difference in timeliness with the hospital registry is indicated by the lag that maximizes R^2. The absenteeism syndrome (green line) preceded the hospital syndrome by 1 week, followed by the GP-based and prescription-based syndromes at +1 week and the syndrome based on mortality and laboratory sub-

Table 3. Pearson correlation coefficients between time series in respiratory pathogen counts, the Netherlands, 1999–2004*†

Pathogen	*S. pneumoniae*	RSV	Influenza A	Influenza B	RV	PIV	Adenovirus	Enterovirus	*Mycoplasma pneumoniae*	*Bordetella pertussis*
S. pneumoniae	1.00	0.35	**0.51**	0.36	NS	0.32	0.32	–0.44	0.21	–0.31
RSV		1.00	0.23	NS	0.30	0.13	0.21	–0.30	0.19	NS
Influenza A			1.00	0.36	NS	0.12	0.24	–0.39	0.16	–0.25
Influenza B				1.00	NS	NS	NS	–0.30	0.25	–0.21
RV					1.00	NS	0.21	NS	NS	NS
PIV						1.00	NS	–0.19	NS	NS
Adenovirus							1.00	–0.21	NS	–0.14
Enterovirus								1.00	–0.15	0.21
M. pneumoniae									1.00	NS
B. pertussis										1.00

S. pneumoniae, Streptococcus pneumoniae*; RSV, respiratory syncytial virus; RV, rhinovirus; PIV, parainfluenza virus; NS, nonsignificant. Correlations ≥0.50 in **boldface; p value ≥0.05.
†Unlagged.

Table 4. All respiratory pathogen counts included as explanatory variables in the regression models, the Netherlands, 1999–2004*†

Syndrome data	RSV	Influenza A	Influenza B	*S. pneumoniae* (residual)	RV	PIV	Adenovirus	Enterovirus	*Mycoplasma pneumoniae*	*Bordetella pertussis*
Absenteeism	2	5	4	2	4	5	–	–	–	
GP	–1	1	2	–1	1	2	–2	–	–3	
Pharmacy	–1	0	2	0	2	5	–2	–	5	–3
Hospitalization	0	2	1	–	–2	3	–	–	–	
Laboratory submissions	–2	0	1	–3	–	2		–	5	
Mortality	–3	1	0	–	–	–	–	–	–	

**S. pneumoniae, Streptococcus pneumoniae*; RSV, respiratory syncytial virus; RV, rhinovirus; PIV, parainfluenza virus; GP, general practice; –, pathogen not included in model.
†The lag time (in weeks) is indicated, that showed optimal fit between syndrome time-series and lagged pathogen counts included in the linear regression model; e.g., according to the model, the trend in hospitalizations precedes the influenza A laboratory counts by 2 weeks.

mission data at +2 weeks after the hospital syndrome (projected on x-axis, Figure 2).

The differences in timeliness between the syndromes and the pathogen surveillance data were reflected by the regression models relating the syndromes to the (positive or negative) lagged pathogens (Table 4). Influenza A and influenza B had lags of 0–5 weeks, which suggests that the registry-syndromes were 0–5 weeks ahead of laboratory counts for these infections. Fluctuations in the time series of respiratory hospitalizations and the laboratory RSV counts seemed to appear in the same week (lag = 0). All other syndromes appeared to be 1–3 weeks later than the RSV counts, except absenteeism, which is 2 weeks earlier. Again, absenteeism seemed to be the earliest syndrome (2–5 weeks earlier than RSV, influenza A, and influenza B), followed by the hospital syndrome (0–2 weeks earlier), the GP-based and prescription-based syndromes (2 weeks earlier until 1 week later), the laboratory submission syndrome (1 week earlier until 2 weeks later), and the mortality syndrome (0–3 weeks later than RSV, influenza A, and influenza B).

Discussion

We explored the potential of 6 Dutch medical registries for respiratory syndromic surveillance. Although several other studies also evaluated routine (medical) data for syndromic surveillance purposes (*22*–*27*), most evaluated only 1 syndrome and correlated this only to influenza data. An exception is Bourgeois et al. (*24*), who validated a respiratory syndrome in relation to diagnoses of several respiratory pathogens in a pediatric population, and Cooper et al. (*27*), who estimated the contribution of specific respiratory pathogens to variations in respiratory syndromes. Both studies concluded that RSV and influenza explain most of the variations in these syndromes, consistent with our findings.

Our study shows that all syndrome data described in this study showed higher levels in winter, which corresponded to the seasonal patterns of RSV, *S. pneumoniae*, and influenza A and B viruses. Linear regression showed that the syndromes can be explained by lagged laboratory counts for respiratory pathogens (up to 86%, highest effect of influenza A, influenza B, and RSV), which indicates their potential usefulness for syndromic surveillance. Timeliness differed, with up to 5 weeks potential gain in early warning by syndromic data, compared with routine laboratory surveillance data.

A limitation of our study is the short duration of our time series, especially for absenteeism and pharmacy data. Therefore, whether our observed associations between syndromes and pathogen counts can be generalized remains unclear.

We relied on laboratory pathogen counts as a proxy for their prevalence and the illness they cause. Changes in test volume over time would result in misclassification bias (as noncausative pathogens will be detected as well). However, such changes are presumably dwarfed by changes during "truly" epidemic elevations of common respiratory pathogens. Additionally, laboratory diagnostics are mostly performed on hospitalized patients, and thus results inadequately reflect activity of pathogens that predominantly cause mild illness.

By adding seasonal terms, we observed that for the absenteeism and, to a lesser extent, the pharmacy registry, the associations between the respiratory syndromes and the pathogen counts might be biased to some extent. For the GP, hospital, laboratory submission, and mortality data,

Table 5. Syndrome variation that can be explained by either the pathogen counts, seasonal terms, or pathogen counts and seasonal terms together*

Syndrome data	Pathogens, %	Pathogens and seasonal terms, %	Seasonal terms, %
Absenteeism	68	80	79
GP	86	89	75
Pharmacy	80	87	77
Hospitalization	84	88	75
Laboratory submissions	61	63	53
Mortality	78	78	52

*Estimated by 3 different R^2 values for each registry: 1) for the syndromes explained by pathogen counts alone; 2) after adding seasonal terms to the pathogen model; and 3) for the syndromes explained by seasonal terms alone (sine and cosine parameters). GP, general practice.

Table 6. Standardized parameter estimates (βs) for all respiratory pathogen counts included as explanatory variables in the regression models, before and after adding seasonal terms to the models*†

Syndrome data	RSV	Influenza A	Influenza B	*S. pneumoniae* (residual)	RV	PIV	Adenovirus	Enterovirus	*Mycoplasma pneumoniae*	*Bordetella pertussis*
Absenteeism	0.31/ (NS)	0.27/ (NS)	0.33/ (NS)	0.28/ 0.12	0.19/ (NS)	0.20/ (NS)	–	–	–	–
GP	0.60/ 0.51	0.32/ 0.32	0.20/ 0.16	0.13/ 0.10	0.07/ (NS)	0.14/ 0.08	0.07/ (NS)	–	0.06/ 0.05	–
Pharmacy	0.51/ 0.54	0.27/ 0.22	0.24/ (NS)	0.25/ 0.11	0.16/ 0.08	0.16/ (NS)	0.08/ (NS)	–	0.12/ (NS)	0.11/ 0.11
Hospitalization	0.60/ 0.44	0.36/ 0.34	0.21/ 0.12	–	0.13/ 0.05	0.09/ (NS)	–	–	–	–
Laboratory submissions	0.49/ 0.47	0.19/ 0.20	0.22/ 0.18	0.28/ 0.22	–	0.17/ 0.08	–	–	0.10/ 0.10	–
Mortality	0.40/ 0.36	0.52/ 0.51	0.24/ 0.24	–	–	–	–	–	–	–

**S. pneumoniae, Streptococcus pneumoniae*; RSV, respiratory syncytial virus; RV, rhinovirus; PIV, parainfluenza virus; GP, general practice; –, pathogen not included in model; NS, the pathogen variable is no longer significant after seasonal terms are added.
†For example, 0.60/0.40 for RSV indicates a standardized β of 0.60 for RSV in the model with only pathogen variables and a β of 0.40 in the same model after adding seasonal terms.

season is probably not an important confounder for the association between the syndromes and pathogens, because including seasonal terms in the models resulted in the same or only slightly higher explained syndrome variance (measured by R^2). Models with seasonal terms alone mostly had lower explained variance than the pathogen models. For the GP and hospital data, some pathogens became insignificant after seasonal terms were added (Table 6) but not those pathogens with the largest effect estimates (RSV, influenza A and B). Therefore, we are confident in concluding that the GP, hospital, laboratory submission, and mortality syndromes do reflect pathogen activity sufficiently for use in syndromic surveillance.

The higher R^2 value of the absenteeism model with seasonal terms alone suggests seasonality of absenteeism caused by several nonrespiratory conditions (*28*,*29*). To some extent, this also applies to the pharmacy syndrome, which includes medications that are not specific for respiratory infections (e.g., antimicrobial drugs). This could be validated in future studies by linking medications to illness. However, for both the absenteeism and pharmacy syndromes, the variation explained by seasonal terms is probably overestimated to some extent because data for only 2 and 3 years were used. Consequently, these time series contained less information on variation between different years than for the other registries, which benefits fitting of a model with several sine and cosine terms.

To our knowledge, laboratory submission data (test requests) have not been evaluated before as a data source for syndromic surveillance. The modest explained variance for the laboratory submissions syndrome could possibly reflect the limited use in our country of laboratory testing algorithms, which leads to substantial differences in diagnostic regimes for patients with similar clinical symptoms. In addition, occasional extra alertness by clinicians can make these data unreliable for surveillance. For instance, an unusual peak was observed in the laboratory submissions syndrome in 1999, after the official announcement of an outbreak of Legionnaires' disease (*30*).

An unexpected increase was also observed in the absenteeism, GP, and pharmacy syndromes, which occurred consistently each year around October (2001–2004). These peaks preceded the syndrome peaks concurring with peaks in influenza A, influenza B, and RSV counts and may be caused by rhinovirus activity—and asthma exacerbations caused by rhinovirus—which usually rises in the fall (*31*–*33*). Rhinovirus might go undetected because GP physicians rarely ask for diagnostics if they suspect a nonbacterial cause for relatively mild respiratory disease. Although

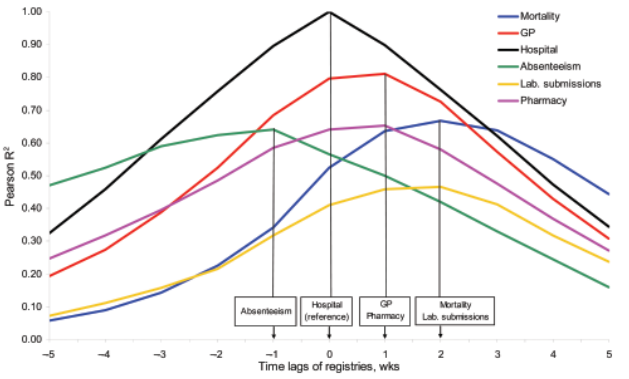

Figure 2. The (maximum) R^2 by the lagged syndromes with the hospital syndrome as a reference. Aggregated by week, univariate Pearson correlation coefficients were calculated of the hospital syndrome and each of the other syndromes. Note that the Pearson correlation coefficients are calculated over different periods for the different registries because not all registries cover the same period (Table 1). Measured by the syndrome lag with the maximized R^2, the timeliness differed between the registries in the following order: absenteeism, hospital, pharmacy/general practice (GP), mortality/laboratory submissions (as projected on the x-axis).

specific asthma diagnoses were excluded from the respiratory syndrome definitions, exacerbations of asthma might affect other respiratory categories in the GP or pharmacy syndrome. This observation illustrates that additional diagnostics are needed for identifying the causes of unexplained respiratory disease elevations. Several novel respiratory pathogens for which diagnostics are not yet widely available have been discovered in recent years, underlining that it is quite possible that "hidden" epidemics occur (*34*–*36*). The extra October peak and several other syndrome elevations above the 95% UCLs in our study may well reflect such hidden epidemics. The fact that these occur is supported by studies showing that many individual syndrome cases cannot be linked to known pathogens. For example, Cooper et al. (*37*), who investigated syndromic signals by using patient self-sampling (at home), could only obtain diagnostic results for 22% of these cases.

For early warning surveillance, timeliness is crucial. Absenteeism data seem to have the best timeliness, but their lack of medical detail complicates interpretation. Unexpectedly, the hospital data reflect respiratory pathogen activity earlier than the GP data. Although in the Netherlands patients are encouraged to consult their GP before going to the hospital, elderly persons, for whom respiratory infections are more likely to cause severe illness, may often go to a hospital directly. Therefore, hospital data may prove to be an earlier marker for respiratory disease than GP data, but this possibility needs further exploration.

An important concern when using syndromic surveillance is that it may generate nonspecific alerts, which, if they happen regularly, would lead to lack of confidence in a syndrome-based surveillance system. Here, we see a clear advantage of using data from multiple registries in parallel so that signal detection can be made more specific by focusing on signals that occur concurrently in >1 data source. To illustrate this we defined every exceeding of the UCLs of the regression models as a "signal," i.e., a syndrome elevation unexplained by known pathogen activity and therefore possibly reflecting activity of underdiagnosed or emerging infectious disease. Over 2002–2003 (the period that all 6 registries were in the study), only 5 "concurrent" signals occurred versus 34 "single" signals over all registries. We did not evaluate whether the syndromes indeed detect outbreaks of infectious diseases earlier than clinical or laboratory pathogen surveillance. Such an evaluation is often performed by testing the ability to detect historical natural outbreaks or simulated outbreaks (*10*,*38*). However, historical natural outbreaks are rare and simulated outbreaks may be unrealistic. Nevertheless, further research into the outbreak detection performance of these syndromes would be worthwhile.

The results of this study suggest that it might be best to combine syndromic data and pathogen counts in a prospective surveillance system. Such surveillance can identify distinct syndrome elevations that cannot be explained by respiratory pathogen activity as indicated by routine laboratory pathogen surveillance.

Conclusion

Overall, the GP, hospital, mortality and, to a lesser extent, laboratory submission syndromes reflect week-to-week fluctuations in the time-series of respiratory pathogens as detected in the laboratory. Registries monitoring trends of these syndromes will therefore most likely reflect illness caused by emerging or underdiagnosed respiratory pathogens as well and therefore are suited for syndromic surveillance. Further research would be required to assess to what extent absenteeism and pharmacy data reflect respiratory illness. Investigating the actual outbreak detection performance of the syndromes in this study would also be worthwhile.

Data from the registries in this study are not yet real-time available, although given modern information technology, this availability is clearly feasible. Our study can help prioritize which type of healthcare data to include in future syndromic real-time surveillance systems.

Acknowledgments

We thank Daan Notermans for his expert opinion on providing syndrome definitions; Mariken van der Lubben for reading and commenting on the manuscript; Statistics Netherlands (CBS, Ingeborg Deerenberg and John Kartopawiro), the Foundation for Pharmaceutical Statistics (SFK, Jan-Dirk Kroon and Fabiënne Griens), the Dutch National Medical Register (LMR, Willem Hoogen Stoevenbeld), and the National Information Network of GPs (LINH, Robert Verheij) for providing data; and the members of the Dutch Working Group on Clinical Virology for collecting and providing weekly positive diagnostic results.

Mr van den Wijngaard is an epidemiologist at the Center for Infectious Disease Control at the National Institute of Public Health and the Environment (RIVM), Bilthoven. His main research interests include the use of healthcare data for infectious disease surveillance and monitoring.

References

1. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis. 2003;9:1197–204.
2. Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. BMC Public Health. 2001;1:9.
3. Fleming DM, Barley MA, Chapman RS. Surveillance of the bioterrorist threat: a primary care response. Commun Dis Public Health. 2004;7:68–72.
4. Miller M, Roche P, Spencer J, Deeble M. Evaluation of Australia's National Notifiable Disease Surveillance System. Commun Dis Intell. 2004;28:311–23.

5. Ohkusa Y, Shigematsu M, Taniguchi K, Okabe N. Experimental surveillance using data on sales of over-the-counter medications—Japan, November 2003–April 2004. MMWR Morb Mortal Wkly Rep. 2005;54(Suppl):47–52.
6. Heffernan R, Mostashari F, Das D, Karpati A, Kuldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis. 2004;10:858–64.
7. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for communitywide influenzalike illness, New York City. J Urban Health. 2003;80:i43–9.
8. Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM. Public health assessment of potential biological terrorism agents. Emerg Infect Dis. 2002;8:225–30.
9. Lazarus R, Kleinman K, Dashevsky I, Adams C, Kludt P, DeMaria A Jr, et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis. 2002;8:753–60.
10. Buckeridge DL. Outbreak detection through automated surveillance: a review of the determinants of detection. J Biomed Inform. 2007;40:370–9.
11. Heijnen ML, Dorigo-Zetsma JW, Bartelds AI, Wilbrink B, Sprenger MJ. Surveillance of respiratory pathogens and influenza-like illnesses in general practices—the Netherlands, winter 1997–98. Euro Surveill. 1999;4:81–4.
12. Hertzberger LI, Huisman J, Wilterdink JB. The global eradication of polio by the year 2000 [in Dutch]. Ned Tijdschr Geneeskd. 1998;142:972–3.
13. Widdowson MA, Bosman A, van Straten E, Tinga M, Chaves S, van Eerden L, et al. Automated, laboratory-based system using the Internet for disease outbreak detection, the Netherlands. Emerg Infect Dis. 2003;9:1046–52.
14. Van den Brandhof WE, Kroes ACM, Bosman A, Peeters MF, Heijnen MLA. Reporting virus diagnostics in the Netherlands: representativeness of the virological weekly reports [in Dutch]. Infectieziekten Bulletin. 2002;13:110–3 [cited 2008 Apr 8]. Available from http://www.rivm.nl/infectieziektenbulletin/bul1304/vir_diagnostiek.html
15. Akaike H. A new look at statistical model identification. IEEE Transactions on Automatic Control. 1974;AU-19:716–23.
16. Avadhanula V, Rodriguez CA, Devincenzo JP, Wang Y, Webby RJ, Ulett GC, et al. Respiratory viruses augment the adhesion of bacterial pathogens to respiratory epithelium in a viral species– and cell type–dependent manner. J Virol. 2006;80:1629–36.
17. Hament JM, Aerts PC, Fleer A, Van Dijk H, Harmsen T, Kimpen JL, et al. Enhanced adherence of *Streptococcus pneumoniae* to human epithelial cells infected with respiratory syncytial virus. Pediatr Res. 2004;55:972–8.
18. Kim PE, Musher DM, Glezen WP, Rodriguez-Barradas MC, Nahm WK, Wright CE. Association of invasive pneumococcal disease with season, atmospheric conditions, air pollution, and the isolation of respiratory viruses. Clin Infect Dis. 1996;22:100–6.
19. Hament JM, Kimpen JL, Fleer A, Wolfs TF. Respiratory viral infection predisposing for bacterial disease: a concise review. FEMS Immunol Med Microbiol. 1999;26:189–95.
20. Heisterkamp SH, Dekkers AL, Heijne JC. Automated detection of infectious disease outbreaks: hierarchical time series models. Stat Med. 2006;25:4179–96.
21. Dekkers ALM, Heisterkamp SH. NPBats, Bayesian statistical instrument for trend detection and time-series modelling [in Dutch]. National Institute for Public Health and the Environment (RIVM), 2004; internal report 550002006 [cited 2008 Apr 15]. Available from http://www.rivm.nl/bibliotheek/rapporten/550002006.pdf
22. Miller B, Kassenborg H, Dunsmuir W, Griffith J, Hadidi M, Nordin JD, et al. Syndromic surveillance for influenzalike illness in ambulatory care network. Emerg Infect Dis. 2004;10:1806–11.
23. Brillman JC, Burr T, Forslund D, Joyce E, Picard R, Umland ET. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance. BMC Med Inform Decis Mak. 2005;5:4.
24. Bourgeois FT, Olson KL, Brownstein JS, McAdam AJ, Mandl KD. Validation of syndromic surveillance for respiratory infections. Ann Emerg Med. 2006;47:265.e1.
25. Smith G, Hippisley-Cox J, Harcourt S, Heaps M, Painter M, Porter A, et al. Developing a national primary care-based early warning system for health protection—a surveillance tool for the future? Analysis of routinely collected data. J Public Health (Oxf). 2007;29:75–82.
26. Vergu E, Grais RF, Sarter H, Fagot JP, Lambert B, Valleron AJ, et al. Medication sales and syndromic surveillance, France. Emerg Infect Dis. 2006;12:416–21.
27. Cooper DL, Smith GE, Edmunds WJ, Joseph C, Gerard E, George RC. The contribution of respiratory pathogens to the seasonality of NHS Direct calls. J Infect. 2007;55:240–8.
28. Fisman DN. Seasonality of infectious diseases. Annu Rev Public Health. 2007;28:127–43.
29. van Rossum CT, Shipley MJ, Hemingway H, Grobbee DE, Mackenbach JP, Marmot MG. Seasonal variation in cause-specific mortality: are there high-risk groups? 25-year follow-up of civil servants from the first Whitehall study. Int J Epidemiol. 2001;30:1109–16.
30. Den Boer JW, Yzerman EP, Schellekens J, Lettinga KD, Boshuizen HC, Van Steenbergen JE, et al. A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. Emerg Infect Dis. 2002;8:37–43.
31. Gwaltney JM Jr, Hendley JO, Simon G, Jordan WS Jr. Rhinovirus infections in an industrial population. I. The occurrence of illness. N Engl J Med. 1966;275:1261–8.
32. Dales RE, Schweitzer I, Toogood JH, Drouin M, Yang W, Dolovich J, et al. Respiratory infections and the autumn increase in asthma morbidity. Eur Respir J. 1996;9:72–7.
33. Van Gageldonk-Lafeber AB, Heijnen ML, Bartelds AI, Peters MF, van der Plas SM, Wilbrink B. A case-control study of acute respiratory tract infection in general practice patients in the Netherlands. Clin Infect Dis. 2005;41:490–7.
34. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. Proc Natl Acad Sci U S A. 2005;102:12891–6.
35. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 2003;348:1967–76.
36. Van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, et al. Identification of a new human coronavirus. Nat Med. 2004;10:368–73.
37. Cooper DL, Smith GE, Chinemana F, Joseph C, Loveridge P, Sebastionpillai P, et al. Linking syndromic surveillance with virological self-sampling. Epidemiol Infect. 2008;136:222–4.
38. Bravata DM, McDonald KM, Smith WM, Rydzak C, Szeto H, Buckeridge DL, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med. 2004;140:910–22.

Address for correspondence: Cees van den Wijngaard, Centre for Infectious Disease Control, Netherlands National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, the Netherlands; email: kees.van.den.wijngaard@rivm.nl

Validation of Syndromic Surveillance for Respiratory Pathogen Activity

Technical Appendix

Detailed Syndrome Definitions for Each Syndrome Data Source

A general respiratory syndrome was defined for each data source (except for the absenteeism data, which contain no medical information; see Table 1). We used the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM*)* codes as selected by the Centers for Disease Control and Prevention (CDC), Atlanta, Georgia, USA (www.bt.cdc.gov/surveillance/syndromedef). To define a respiratory syndrome, we selected both the codes for general respiratory symptoms and diagnoses (category 1 in CDC list) and the codes for specific respiratory biologic agent diagnoses (category 3 in CDC list). For the hospital data (see Table 1), we used these syndrome codes with some minor adaptations for the Dutch version of ICD-9-CM. For the mortality data (see Table 2) the ICD-9-CM codes were converted into ICD 10th revision (ICD-10) codes by using the World Health Organization ICD-9/ICD-10 translation list and expert opinion, if necessary (ICD-9/ICD-10 Translator; see www.who.int/classifications/en). For the GP consultation data (see Table 3), International Classification of Primary Care (ICPC) codes were included in a respiratory syndrome by expert opinion, guided by the CDC respiratory syndrome case definition.

For a respiratory syndrome definition based on the pharmacy data, we used Anatomical Therapeutic Chemical Classification System (ATC) codes of medications that experts considered indicative for respiratory infectious disease complaints. Of those, we included only ATC-5 codes that had higher levels in winter. See Table 4 for the specific included ATC-5 codes.

For a respiratory syndrome definition based on the laboratory submissions data, we included all submissions for specific diagnostics that are known to be of respiratory cause: 1) all submissions for microbiologic diagnostic tests on respiratory materials (sputum, bronchoalveolar lavage, pleural liquid); 2) all submissions for serology on known specific respiratory pathogens

(see list of serologic tests in Table 5); 3) all submissions for *Legionella* spp. or *Streptococcus pneumoniae* antigen tests on urine.

For all data types we assumed that in a prospective setting real-time syndrome-classification would be feasible (on date of consultation/hospitalization/death/submission/ dispense).

Table 1. ICD-9-CM codes for the respiratory syndrome in hospital data*

ICD-9-CM code	Description
020.3	Primary pneumonic plague
020.4	Secondary pneumonic plague
020.5	Pneumonic plague not otherwise specified
021.2	Pulmonary tularemia
022.1	Pulmonary anthrax
031.0	Mycobacteria, pulmonary
031.8	Other specified mycobacterial diseases
031.9	Mycobacteria diseases/unspecified
032.0	Faucial diphtheria
032.1	Nasopharynx diphtheria
032.2	Anterior nasal diphtheria
032.3	Laryngeal diphtheria
032.89	Diphtheria not elsewhere classified
032.9	Diphtheria not otherwise specified
033.0	*Bordetella pertussis*
033.1	*Bordetella parapertussis*
033.8	Whooping cough not elsewhere classified
033.9	Whooping cough (unspecified organism)
034.0	Streptococcal sore throat
055.1	Postmeasles pneumonia
055.2	Postmeasles otitis media
073.0	Ornithosis, with pneumonia
073.7	Ornithosis, with other specified complication
073.8	Ornithosis, with unspecified complication
073.9	Ornithosis, unspecified
079.0	Adenovirus infection not otherwise specified
079.1	Echovirus infection not otherwise specified nos.
079.2	Coxsackie virus
079.3	Rhinovirus infection not otherwise specified
079.8	Viral infection in conditions classified elsewhere and of unspecified site
098.6	Gonoccocal, infection of pharynx
114.0	Primary coccidioidomycosis (lung)

114.5	Pulmonary coccidioidomycosis, unspecified
114.9	Coccidioidomycosis not otherwise specified
115.00	Histoplasmosis, without mention of manifestation
115.05	*Histoplasma capsulatum* pneumonia
115.09	*Histoplasma capsulatum* not elsewhere classified
115.10	*Histoplasma duboisii* not otherwise specified
115.15	*Histoplasma duboisii* pneumonia
115.90	Histoplasmosis, without manifestation
115.95	Histoplasmosis pneumonia
115.99	Histoplasmosis not elsewhere classified
116.0	Blastomycosis
116.1	Paracoccidioidomycosis
117.1	Sporotrichosis
117.3	Pulmonary aspergillosis
117.5	Cryptococcosis
130.4	*Toxoplasma* pneumonitis
136.3	Pneumocystosis
460	Nasopharyngitis, acute
462	Pharyngitis, acute not otherwise specified
463	Tonsillitis, acute
464.0	Acute laryngitis
464.10	Tracheitis without obstruction
464.11	Acute tracheitis with obstruction
464.20	Laryngotracheitis without obstruction
464.21	Acute laryngotracheitis with obstruction
464.30	Epiglottitis acute without obstruction
464.31	Acute epiglottitis with obstruction
464.4	Croup
465.0	Laryngopharyngitis, acute
465.8	Upper respiratory infection, other multiple sites
465.9	Upper respiratory infection, acute not otherwise specified
466.0	Bronchitis acute
466.1	Acute bronchiolitis
478.9	Respiratory tract disease
480.0	Adenoviral pneumonia
480.1	Pneumonia due to respiratory syncytial virus
480.2	Parinfluenza viral pneumonia
480.8	Viral pneumonia not elsewhere classified
480.9	Pneumonia, viral
481	Pneumococcal pneumonia (lobar)
482.0	Pneumonia due to *Klebsiella pneumoniae*
482.1	Pneumonia due to *Pseudomonas*
482.2	*Haemophilus influenzae* pneumonia
482.3	Pneumonia due to *Streptococcus*

482.4	Pneumonia due to *Staphylococcus*
482.8	Pneumonia due to bacteria not elsewhere classified
482.9	Pneumonia due to bacteria not otherwise specified
483	Pneumonia due to organism not elsewhere classified
484.1	Pneumonia due to cytomegalic inclusion disease
484.3	Pneumonia in whooping cough
484.5	Pneumonia in anthrax
484.6	Pneumonia in aspergillosis
484.7	Pneumonia in other systemic mycoses
484.8	Pneumonia in infection disease not elsewhere classified
485	Bronchopneumonia organism unspecified
486	Pneumonia, organism not otherwise specified
487.0	Influenza with pneumonia
487.1	Influenza with other respiratory manifestations
487.8	Influenza with other manifestations
490	Bronchitis not otherwise specified
511.0	Pleurisy without mention of effusion or current tuberculosis
511.1	Pleurisy with effusion, with mention of a bacterial cause other than tuberculosis
511.8	Hemothorax
513.0	Abscess lung
513.1	Abscess of mediastinum
518.4	Edema lung acute not otherwise specified
518.8	Other diseases of lung not otherwise classified
519.2	Mediastinitis
519.3	Mediastinum, diseases not elsewhere classified
769	Respiratory distress syndrome
786.00	Respiratory abnormality
786.09	Other specified respiratory abnormality
786.1	Stridor
786.2	Cough
786.3	Hemoptysis
786.52	Painful respiration/pleurodynia
799.1	Respiratory arrest

*ICD-9-CM, International Classification of Diseases, 9th Revision, Clinical Modification.

Table 2. ICD-10 codes for the respiratory syndrome in mortality data*

ICD-10 code	Description
A202	Pneumonic plague
A212	Pulmonary tularemia
A221	Pulmonary anthrax
A310	Pulmonary mycobacterial infection
A318	Other mycobacterial infections
A319	Mycobacterial infection, unspecified
A360	Pharyngeal diphtheria

A361	Nasopharyngeal diphtheria
A362	Laryngeal diphtheria
A368	Other diphtheria
A369	Diphtheria, unspecified
A370	Whooping cough due to *Bordetella pertussis*
A371	Whooping cough due to *Bordetella parapertussis*
A378	Whooping cough due to other *Bordetella* species
A379	Whooping cough, unspecified
A481	Legionnaires' disease
A545	Gonococcal pharyngitis
A70	*Chlamydia psittaci* infection
B012	Varicella pneumonia (J17.1*)
B052	Measles complicated by pneumonia (J17.1*)
B053	Measles complicated by otitis media (H67.1*)
B340	Adenovirus infection, unspecified
B341	Enterovirus infection, unspecified
B342	Coronavirus infection, unspecified
B348	Other viral infections of unspecified site
B380	Acute pulmonary coccidioidomycosis
B382	Pulmonary coccidioidomycosis, unspecified
B389	Coccidioidomycosis, unspecified
B390	Acute pulmonary histoplasmosis capsulatum
B392	Pulmonary histoplasmosis capsulatum, unspecified
B393	Disseminated histoplasmosis capsulatum
B394	Histoplasmosis capsulatum, unspecified
B395	Histoplasmosis duboisii
B399	Histoplasmosis, unspecified
B400	Acute pulmonary blastomycosis
B402	Pulmonary blastomycosis, unspecified
B407	Disseminated blastomycosis
B408	Other forms of blastomycosis
B409	Blastomycosis, unspecified
B410	Pulmonary paracoccidioidomycosis
B417	Disseminated paracoccidioidomycosis
B418	Other forms of paracoccidioidomycosis
B419	Paracoccidioidomycosis, unspecified
B420	Pulmonary sporotrichosis (J99.8*)
B427	Disseminated sporotrichosis
B428	Other forms of sporotrichosis
B429	Sporotrichosis, unspecified
B440	Invasive pulmonary aspergillosis
B441	Other pulmonary aspergillosis
B442	Tonsillar aspergillosis
B447	Disseminated aspergillosis

B448	Other forms of aspergillosis
B449	Aspergillosis, unspecified
B450	Pulmonary cryptococcosis
B457	Disseminated cryptococcosis
B458	Other forms of cryptococcosis
B459	Cryptococcosis, unspecified
B583	Pulmonary toxoplasmosis (J17.3*)
B59	Pneumocystosis
B970	Adenovirus as the cause of diseases classified to other chapters
B971	Enterovirus as the cause of diseases classified to other chapters
B972	Coronavirus as the cause of diseases classified to other chapters
B974	Respiratory syncytial virus as the cause of diseases classified to other chapters
B978	Other viral agents as the cause of diseases classified to other chapters
G473	Sleep apnea
J00	Acute nasopharyngitis (common cold)
J020	Streptococcal pharyngitis
J028	Acute pharyngitis due to other specified organisms
J029	Acute pharyngitis, unspecified
J030	Streptococcal tonsillitis
J038	Acute tonsillitis due to other specified organisms
J039	Acute tonsillitis, unspecified
J040	Acute laryngitis
J041	Acute tracheitis
J042	Acute laryngotracheitis
J050	Acute obstructive laryngitis (croup)
J051	Acute epiglottitis
J060	Acute laryngopharyngitis
J068	Other acute upper respiratory infections of multiple sites
J069	Acute upper respiratory infection, unspecified
J100	Influenza with pneumonia, influenza virus identified
J101	Influenza with other respiratory manifestations, influenza virus identified
J108	Influenza with other manifestations, influenza virus identified
J110	Influenza with pneumonia, virus not identified
J111	Influenza with other respiratory manifestations, virus not identified
J118	Influenza with other manifestations, virus not identified
J120	Adenoviral pneumonia
J121	Respiratory syncytial virus pneumonia
J122	Parainfluenza virus pneumonia
J128	Other viral pneumonia
J129	Viral pneumonia, unspecified
J13	Pneumonia due to *Streptococcus pneumoniae*
J14	Pneumonia due to *Haemophilus influenzae*
J150	Pneumonia due to *Klebsiella pneumoniae*
J151	Pneumonia due to *Pseudomonas*

J152	Pneumonia due to *Staphylococcus*
J153	Pneumonia due to *Streptococcus*, group B
J154	Pneumonia due to other streptococci
J155	Pneumonia due to *Escherichia coli*
J156	Pneumonia due to other aerobic Gram-negative bacteria
J157	Pneumonia due to *Mycoplasma pneumoniae*
J158	Other bacterial pneumonia
J159	Bacterial pneumonia, unspecified
J160	Chlamydial pneumonia
J168	Pneumonia due to other specified infectious organisms
J170	Pneumonia in bacterial diseases classified elsewhere
J171	Pneumonia in viral diseases classified elsewhere
J172	Pneumonia in mycoses
J173	Pneumonia in parasitic diseases
J178	Pneumonia in other diseases classified elsewhere
J180	Bronchopneumonia, unspecified
J182	Hypostatic pneumonia, unspecified
J188	Other pneumonia, organism unspecified
J189	Pneumonia, unspecified
J200	Acute bronchitis due to *Mycoplasma pneumoniae*
J201	Acute bronchitis due to *Haemophilus influenzae*
J202	Acute bronchitis due to streptococcus
J203	Acute bronchitis due to coxsackievirus
J204	Acute bronchitis due to parainfluenza virus
J205	Acute bronchitis due to respiratory syncytial virus
J206	Acute bronchitis due to rhinovirus
J207	Acute bronchitis due to echovirus
J208	Acute bronchitis due to other specified organisms
J209	Acute bronchitis, unspecified
J210	Acute bronchiolitis due to respiratory syncytial virus
J218	Acute bronchiolitis due to other specified organisms
J219	Acute bronchiolitis, unspecified
J22	Unspecified acute lower respiratory infection
J398	Other specified diseases of upper respiratory tract
J40	Bronchitis, not specified as acute or chronic
J850	Gangrene and necrosis of lung
J851	Abscess of lung with pneumonia
J852	Abscess of lung without pneumonia
J853	Abscess of mediastinum
J942	Hemothorax
J949	Pleural condition, unspecified
J960	Acute respiratory failure
J969	Respiratory failure, unspecified
J985	Diseases of mediastinum, not elsewhere classified

J998	Respiratory disorders in other diseases classified elsewhere
P220	Respiratory distress syndrome of newborn
R042	Hemoptysis
R049	Hemorrhage from respiratory passages, unspecified
R05	Cough
R061	Stridor
R063	Periodic breathing
R064	Hyperventilation
R065	Mouth breathing
R068	Other and unspecified abnormalities of breathing
R071	Chest pain on breathing
R091	Pleurisy
R092	Respiratory arrest

*ICD-10, International Classification of Diseases, 10th Revision.

Table 3. ICPC codes for the respiratory syndrome in general practice consultations data*

ICPC codes	Description
H71	Acute otitis media/myringitis
L04	Chest symptom/complaint
R01	Pain respiratory system
R02	Shortness of breath/dyspnea
R03	Wheezing
R04	Breathing problem, other
R05	Cough
R07	Sneezing/nasal congestion
R21	Throat symptom/complaint
R24	Hemoptysis
R29	Respiratory symptom/complaint, other
R71	Whooping cough
R74	Upper respiratory infection, acute
R75	Sinusitis acute/chronic
R76	Tonsillitis, acute
R77	Laryngitis/tracheitis acute
R78	Acute bronchitis/bronchiolitis
R80	Influenza
R81	Pneumonia
R82	Pleurisy/pleural effusion
R83	Respiratory infection, other
R93	Pleural effusion not otherwise specified
R99	Respiratory disease, other

*ICPC, International Classification of Primary Care.

Table 4. ATC level 5 codes for the respiratory syndrome in pharmacy data*

ATC-5 code	Description
J01AA	Tetracyclines
J01CA	Penicillins with extended spectrum
J01CR	Combinations of penicillins, including β-lactamase inhibitors
J01FA	Macrolides
R05CA	Expectorants
R05DA	Opium alkaloids and derivatives
R06AD	Phenothiazine derivatives

*ATC, Anatomical Therapeutic Chemical Classification System.

Table 5. Serologic test subjects included in the respiratory syndrome for laboratory submissions (see information on other included tests in text)

Serologic tests performed on
Adenovirus 2
Adenovirus
Antibodies to adenovirus
Antibodies to *Aspergillus fumigatus*
Antibodies to *Aspergillus* species
Antibodies to *Chlamydia pneumoniae*
Antibodies to *Chlamydia psittaci*
Antibodies to *Chlamydia* species
Antibodies to coronavirus
Antibodies to *Corynebacterium diphtheriae*
Antibodies to influenza A virus
Antibodies to influenza B virus
Antibodies to *Legionella*
Antibodies to *Legionella pneumophila*
Antibodies to *Legionella pneumophila* serogroup 1
Antibodies to *Mycoplasma pneumoniae*
Antibodies to parainfluenza 1 virus
Antibodies to parainfluenza 2 virus
Antibodies to parainfluenza 3 virus
Antibodies to parainfluenza virus
Antibodies to respiratory syncytial virus
Antibodies to *Streptococcus pneumoniae*
Antigen *Aspergillus fumigatus*
Antigen *Aspergillus* species
IgA *Chlamydia pneumoniae*
IgA *Chlamydia* species
IgA *Mycoplasma pneumoniae*
IgG adenovirus
IgG *Leptospira*
IgG *Aspergillus fumigatus*
IgG *Chlamydia pneumoniae*
IgG *Chlamydia psittaci*
IgG *Chlamydia* species
IgG influenza virus A

IgG influenza virus B
IgG *Legionella pneumophila*
IgG *Legionella* species
IgG *Mycoplasma pneumoniae*
IgG parainfluenza 1 virus
IgG parainfluenza 2 virus
IgG parainfluenza 3 virus
IgG respiratory syncytial virus
IgG *Streptococcus pneumoniae*
IgM influenza virus A
IgM *Chlamydia psittaci*
IgM *Chlamydia* species
IgM influenza B virus
IgM *Legionella pneumophila*
IgM *Legionella* species
IgM *Mycoplasma pneumoniae*
IgM *Mycoplasma* species
IgM parainfluenza 1 virus
IgM parainfluenza 2 virus
IgM parainfluenza 3 virus

*IG, immunoglobulin.

Details on the Regression Model Variables

We constructed a multiple linear regression model:

$S_t = b_0 + b_1P_{A,t+x} + b_2P_{B,t+y}+\ldots+R_{,t}$

S = level of a respiratory syndrome

t = time in weeks

$P_{A/B/etc}$ = lagged respiratory pathogens detected in the laboratory

x/y/etc = lag time in weeks, for shifting the pathogen time series over a range of -5 up to +5 weeks.

R = residual of the model

A forward stepwise regression approach was used, each step selecting the lagged pathogen that contributed most to the model fit (assessed with Akaike's information criterion). Each pathogen was included in the model only once and only if it contributed significantly ($p<0.05$). Negative associations were excluded to avoid biologically implausible associations in the models between

the pathogens and the syndromes (e.g., negative associations between enteroviruses, which peak in summer, and respiratory syndromes, which peak in winter). We checked for significant autocorrelation in the residual of the models.

To investigate whether seasonal variation could be a confounder for the association between pathogens and syndromes we then calculated 3 R^2 values for the models: 1) with only pathogen variables, 2) after adding seasonal terms (sine(k2πweek/52) and cosine(k2πweek/52), k = 1, 2, 3), and 3) with only seasonal terms. We calculated the standardized parameter estimates as well, before and after adding seasonal terms. The standardized parameter estimates are the beta values that result when all variables are standardized to a mean of 0 and a variance of 1. These estimates are computed by multiplying the original estimates by the standard deviation of the regressor (independent) variable and then dividing by the standard deviation of the dependent variable.