Is This an Outbreak?

A retrospective evaluation of syndromic surveillance for emerging infectious disease

Is This an Outbreak?

A retrospective evaluation of syndromic surveillance for emerging infectious disease

Is dit een uitbraak?

Een retrospectieve evaluatie van syndroomsurveillance voor nieuw opduikende infectieziekten

Proefschrift
ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. H.G. Schmidt
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 9 december 2010 om 13.30 uur

door
Cornelis Carolus van den Wijngaard

geboren te Sittard

ERASMUS UNIVERSITEIT ROTTERDAM

ISBN 978 90 6464 426 9

Layout: GVO drukkers & vormgevers B.V. | Ponsen & Looijen
Printed by: GVO drukkers & vormgevers B.V. | Ponsen & Looijen

Reprints were made with permission of The American Public Health Association and Cambridge University Press.

Publication of this thesis was financially supported by the National Institute for Public Health and the Environment (RIVM) and Erasmus University Rotterdam (EUR).

Promotiecommissie

Promotoren:	Prof.dr. M.P.G. Koopmans Prof.dr. N.J.D. Nagelkerke
Overige leden:	Prof.dr. A. Osterhaus Prof.dr. J.H. Richardus Prof.dr. M.C.J.M. Sturkenboom
Copromotor:	Dr. W. van Pelt

Contents

Chapter **1**

General Introduction

General Introduction

Background

In the last decade, worldwide several major infectious disease events occurred — like the anthrax attacks in the USA in 2001, the SARS epidemic in 2003 and the 2009 influenza pandemic. As a result, public-health authorities worldwide have acknowledged the need for improved surveillance for emerging infectious diseases, as early detection and control may well mitigate the impact of emerging outbreaks. For instance, the SARS epidemic could have caused a major pandemic with millions of deaths if it had not been contained by public health measures. Still, the SARS epidemic was only recognized several months after its emergence, and in the end it infected 8096 patients in 27 countries and caused 774 deaths before it could be controlled [1,2]. SARS, and most other pathogens that are considered a high threat for public health, cause symptoms that are common in clinical practice, like pneumonia (e.g. *B. anthracis,* SARS or pandemic influenza*),* gastro-enteritis (e.g. *Vibrio cholerae*) or neurological disease (e.g. West Nile virus) [3,4]. That is why traditional outbreak detection based on astute clinicians and laboratory diagnoses can have blind spots for such emerging diseases, because patients reporting with such common symptoms might not alarm clinicians — especially since individual clinicians may only see one or a few of these "new" cases — and uncommon pathogens can remain undetected by the laboratory.

To reveal such blind spots, many countries have implemented so called syndromic surveillance systems that aim to capture infectious disease events earlier and more completely than traditional surveillance; they do this by monitoring new health indicators, such as basic symptom information or clinical diagnoses, rather than positive laboratory results for specific pathogens [5-14]. Surveillance of symptom based data has been used since decades for surveillance of polio (acute flaccid paralysis) and influenza (influenza-like illness) [15,16], but the increasing availability of electronic health-care data with information on specific morbidity in time, has made large-scale real-time monitoring of syndromes possible. Syndromic surveillance was initially developed for early-warning detection of bioterrorism attacks, but is also used for early detection of naturally occurring (local) outbreaks, following the size and spread of ongoing outbreaks, monitoring disease trends in the general population, and providing reassurance that an outbreak has not occurred [5,6]. At the same time, in light of limited resources for public health, there has been an ongoing debate about the actual added value of syndromic surveillance, with particular concerns about its specificity [17-20].

Because little validation had been done to address these concerns[18,20], we chose to evaluate the potential value of syndromic surveillance for infectious disease surveillance and control, before starting any implementation. This thesis is the result of that evaluation project.

The concept of syndromic surveillance

Syndromic data

There is no official definition of syndromic surveillance, but systems designated as such monitor all kinds of pre- or non-diagnostic data in order to capture (infectious) disease dynamics. In this context, a syndrome can be defined as a combination of complaints and symptoms of a patient which could indicate an infection with certain

pathogens [9,21-24]. Figure 1 shows that in all disease phases, data can be generated for syndromic surveillance. This ranges from data reflecting mild morbidity not even requiring medical treatment, to data reflecting more severe morbidity (requiring hospitalization) or even death. In the prodromal phase, elevations of absenteeism and over-the-counter sales of medications could indicate early disease elevations. In the clinical phase, work-diagnoses, (acute) chief complaints, prescription medications or emergency-department diagnoses can be used for syndromic surveillance. Trends in the number of specimens submitted to diagnostic laboratories — before the test results become available — could indicate disease elevations as well. Even in the final phase of disease, cause-of-death or crude (overall) mortality data can be used for syndromic surveillance purposes. Traditional surveillance, on the other hand, is mainly limited to diagnoses based on laboratory detection of pathogens (Figure 1), followed by notification.

Analyses and Response

A syndromic surveillance system that is used for early warning, is supposed to generate a signal if a certain syndrome shows an unexpected elevation [5,25]. For signaling such elevations in time and/or space-time, statistical algorithms can be used to compare the difference between the number of observed and expected cases, often derived from historical data [5,6]. Obviously, in order to have added value for public health, syndromic surveillance signals have to be translated in public-health responses such as active case finding, treatment and/or quarantine.

Performance evaluation

To evaluate the added value of syndromic surveillance, it would make sense to assess its performance for different systems and disease events in a standardized way. However, one should keep in mind that quantitative performance measures like the sensitivity, specificity and timeliness of outbreak detection are difficult to standardize, because detection of disease events depends on many sometimes unpredictable factors; these include data types and sources, syndrome definitions, detection algorithms, response actions, surveillance populations and types of disease event [25-27]. In addition, there is a lack of data on real (major) disease events to use as "gold standard", whereas simulated test outbreaks may not reflect the diversity and unpredictability of real-life disease events [25,28]. An additional problem is that many existing syndromic surveillance systems have limitations like low data quality or coverage, making these systems unfit for evaluating the potential performance and thus added value of syndromic surveillance.

Objective, approach, specific research questions and outline of the PhD-thesis

The objective of this PhD-thesis is to evaluate the added value of syndromic surveillance for infectious disease surveillance and control, and to make recommendations for its possible implementation in the Netherlands. We chose to evaluate a range of syndromic data types from existing Dutch medical registries — all retrospective but with high-quality data and preferably high coverage. Table 1 lists the six registries included in the study, with data on work-absenteeism, General Practitioner (GP) consultations, pharmacy prescriptions, laboratory submissions, hospital diagnoses and mortality. Data was available from 1999-2009 or parts of this period.

Rather than estimating performance parameters like sensitivity and specificity for detecting fictitious major

disease events, we first assessed whether known infectious disease dynamics and outbreaks during the study period were reflected in these data sources, assuming that good reflection of known dynamics and outbreaks implied good reflection of emerging pathogen activity as well. Therefore we selected respiratory, gastro-enteritis and/or neurological syndromes for analysis. These syndromes can be expected to reflect the clinical presentations of both high-threat and common pathogens [3,4,21]. 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 recognize by clinicians. See the appendix of chapter 1 for details on the syndrome definitions for each registry.

The specific research questions addressed were:

1. What syndromic data types track known dynamics of infectious diseases in the general population, and thus will also likely reflect emerging pathogen activity (chapter 2)?
2. Can syndromic surveillance improve the monitoring of disease burden and/or detect shifts in virulence of common pathogens (chapter 3)?
3. Can syndromic surveillance detect known local outbreaks with a limited number of signals in time, independent of laboratory detection of the causative pathogens (as an indication for sensitivity and specificity, chapters 4 and 5)?

In chapter 6 we give a perspective on the added value and possible applications of syndromic surveillance and make recommendations for implementation, based upon the results of our project.

In chapter 7 we further elaborate on the implications of this thesis' results for syndromic surveillance system requirements regarding data, analyses and response, in light of the current literature. We also discuss future challenges and possibilities of syndromic surveillance, and conclude upon the added value of syndromic surveillance for an improved infectious disease surveillance.

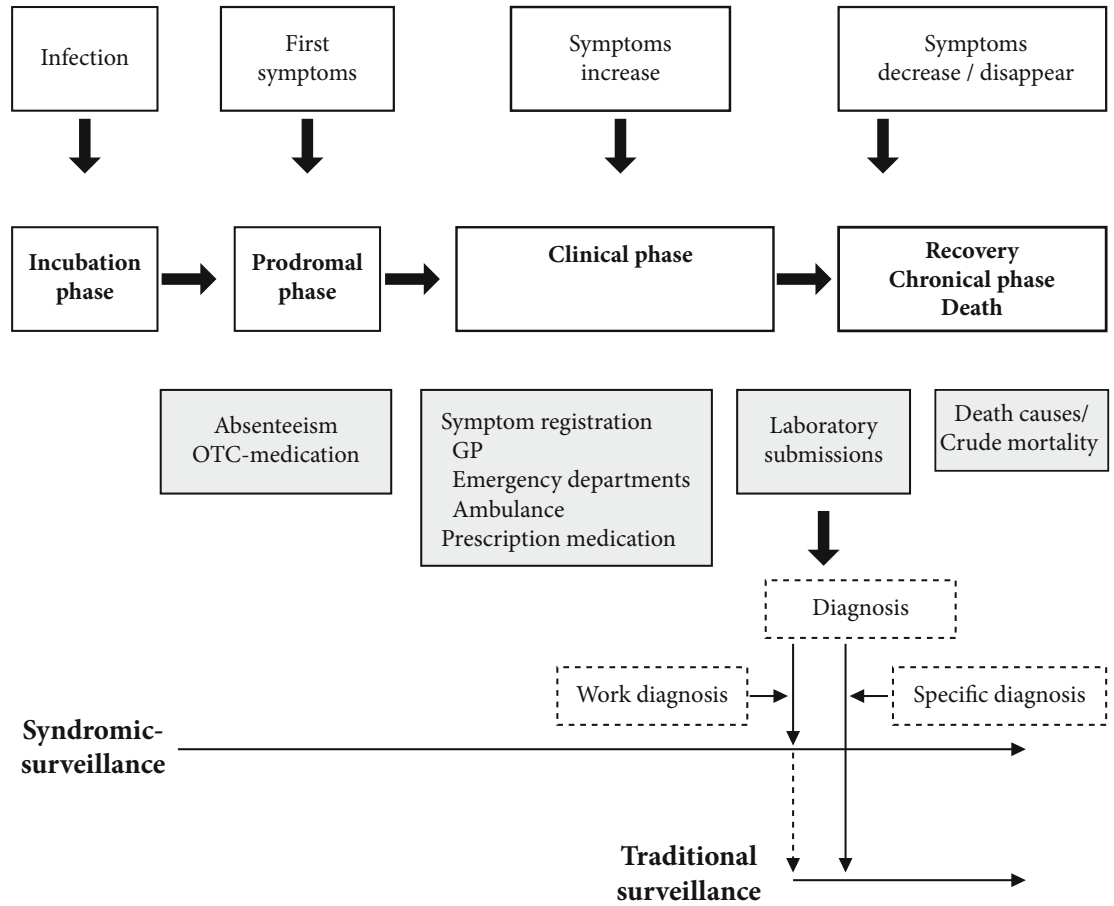


Figure 1. Phases in time of infectious diseases in relation to possible data sources for syndromic and traditional surveillance.

Table 1. Syndromic data registries included in syndromic surveillance evaluation study.

Data type	Period	Coverage (16.3 million pop)	Syndrome definitions (detailed syndrome definitions and codes available on request)	International code system	Analyzed data	Registry
Absenteeism*	2002-2003	80% (of working population, 8 million)	Employers reported sick, no further medical information	-	Sick leave reports of employers	Statistics Netherlands (CBS) (http://www.cbs.nl)
General Practice consultations	2001-2004	1-2%	Symptoms and diagnoses indicating infectious disease	ICPC (International Classification of Primary Care)	Symptoms and diagnoses recorded in practice or telephone consultations, and home visits	Netherlands Information Network of General Practice (LINH) (http://www.nivel.nl/linh/)
Pharmacy dispensations	2001-2003	85%	Prescribed medications indicative for infectious disease	ATC (Anatomical Therapeutic Chemical Classification System)	Prescription medications dispensed in Dutch pharmacies, coded according to the WHO Anatomical Therapeutic Chemical (ATC) classification	Foundation for Pharmaceutical Statistics (http://www.sfk.nl)
Hospitalizations	1999-2007	99%	General symptoms/ diagnoses(cdc1) *and* specific biologic agent diagnoses (cdc3)	ICD-9-CM (International Classification of Diseases, 9th revision, Clinical Modification)	Discharge and secondary diagnoses and date of hospitalization	Dutch National Medical Register (LMR, http://www.dutchhospitaldata.nl/)
Laboratory submissions* (negative and positive results)	2001-2004	16%	Submissions for microbiological diagnostic tests on materials; *and* all submissions for serology on known specific pathogens	-	Lab.submission requests for diagnostic testing	National Infectious Diseases Information System (ISIS)
Mortality	1999-2004; crude (overall) mortality 1999-2009	100%	General symptoms/ diagnoses *and* specific biologic agent diagnoses	ICD-10 (Internat. Classification of Diseases, 10th revision)	Date of death, primary cause of death, complicating and other additional causes of death	Statistics Netherlands (CBS) (http://www.cbs.nl)

*The Lab-requests registry (ISIS-labs) and the absenteeism registry stopped to exist during the project.

Reference List

1. Lingappa JR, McDonald LC, Simone P, Parashar UD. Wrestling SARS from uncertainty. Emerg Infect Dis 2004;10:167-70.
2. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. WHO, 2004. Available at http://www.who.int/csr/sars/country/table2004_04_21/en/index.html, accessed at May 28th 2010.
3. Wagner MM, Dato V, Dowling JN, Allswede M. Representative threats for research in public health surveillance. J Biomed Inform 2003;36:177-88.
4. 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.
5. Henning KJ. Syndromic Surveillance. In Institute of Medicine 2003. Microbial Threats to Health: Emergence, Detection, and Response. Ed. Smolinski MS, Hamburg MA, Lederberg J. The National Academies Press, Washington, DC. 281-312. Available at http://books.nap.edu/openbook.php?record_id=10636&page=281#, accessed at September 7th 2010.
6. Henning KJ. What is syndromic surveillance? MMWR Morb Mortal Wkly Rep 2004;53 Suppl:5-11.
7. Muscatello DJ, Churches T, Kaldor J, Zheng W, Chiu C, Correll P, et al. An automated, broad-based, near real-time public health surveillance system using presentations to hospital Emergency Departments in New South Wales, Australia. BMC Public Health 2005;5:141.
8. Cooper DL, Verlander NQ, Smith GE, Charlett A, Gerard E, Willocks L, et al. Can syndromic surveillance data detect local outbreaks of communicable disease? A model using a historical cryptosporidiosis outbreak. Epidemiol Infect 2006;134:13-20.
9. 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.
10. Van-Dijk A, Aramini J, Edge G, Moore KM. Real-time surveillance for respiratory disease outbreaks, Ontario, Canada. Emerg Infect Dis 2009;15:799-801.
11. Bellika JG, Hasvold T, Hartvigsena G. Propagation of program control: A tool for distributed disease surveillance. Int J Med Inform 2007;76:313-29.
12. Fleming DM, Barley MA, Chapman RS. Surveillance of the bioterrorist threat: a primary care response. Commun Dis Public Health 2004;7:68-72.
13. Flamand C, Larrieu S, Couvy F, Jouves B, Josseran L, Filleul L. Validation of a syndromic surveillance system using a general practitioner house calls network, Bordeaux, France. Euro Surveill 2008;13:pii=18905.
14. Wu TS, Shih FY, Yen MY, Wu JS, Lu SW, Chang KC, et al. Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. BMC Public Health 2008;8:18.
15. Hertzberger LI, Huisman J, Wilterdink JB. [The global eradication of polio by the year 2000]. Ned Tijdschr Geneeskd 1998;142:972-3.
16. 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.
17. Balter S, Weiss D, Hanson H, Reddy V, Das D, Heffernan R. Three years of emergency department gastrointestinal syndromic surveillance in New York City: what have we found? MMWR Morb Mortal Wkly Rep 2005;54 Suppl:175-80.
18. Hope K, Durrheim DN, d'Espaignet ET, Dalton C. Syndromic Surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 2006;60:374-5.
19. Buckeridge DL, Owens DK, Switzer P, Frank J, Musen MA. Evaluating detection of an inhalational anthrax outbreak. Emerg Infect Dis 2006;12:1942-9.
20. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterror 2003;1:77-81.
21. Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef/word/syndromedefinitions.doc, accessed at May 28th 2010.
22. Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health 2003;80:i32-42.
23. Ivanov O, Wagner MM, Chapman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. Proc AMIA Symp 2002;345-9.
24. Zelicoff A, Brillman J, Forslund DW, George JE, Zink S, Koenig S, et al. The Rapid Syndrome Validation Project (RSVP). Proc AMIA Symp 2001;771-5.

25. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recomm Rep 2004;53:1-11.
26. Buckeridge DL. Outbreak detection through automated surveillance: A review of the determinants of detection. J Biomed Inform 2007;40:370-9.
27. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 2003;9:1197-204.
28. Watkins RE, Eagleson S, Hall RG, Dailey L, Plant AJ. Approaches to the evaluation of outbreak detection methods. BMC Public Health 2006;6:263.

Chapter **2**

Validation of Syndromic Surveillance for Respiratory Pathogen Activity

Emerg Infect Dis 2008;14:917-25

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.

*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

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 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 registries 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).

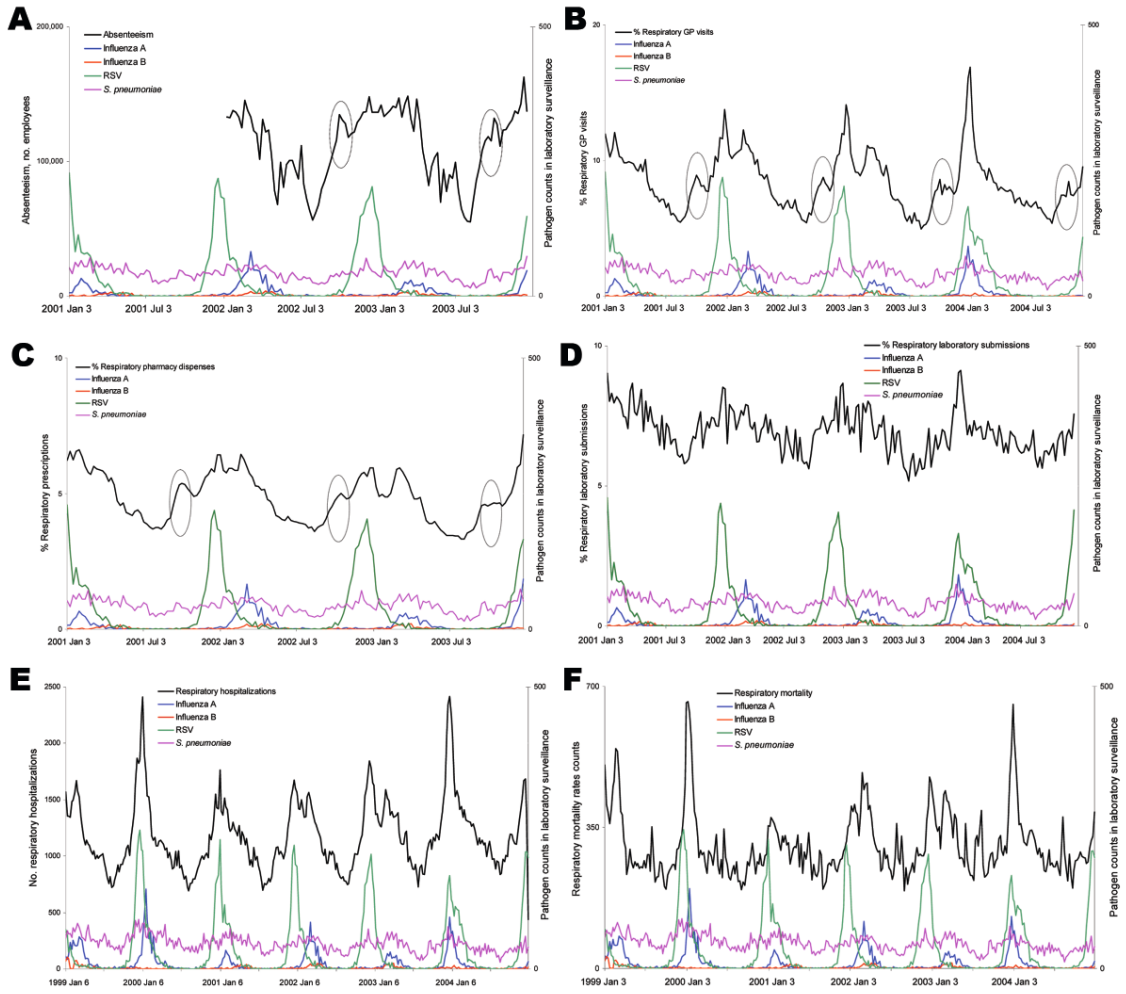


Figure 1. Respiratory syndrome time series and laboratory pathogen counts in the Netherlands. Respiratory syndromes were defined for the 6 registries defined in Table 1: A) absenteeism, B) general practice (GP) consultations, C) pharmacy, D) laboratory submissions, E) hospitalizations, and F) mortality 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.

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

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.

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

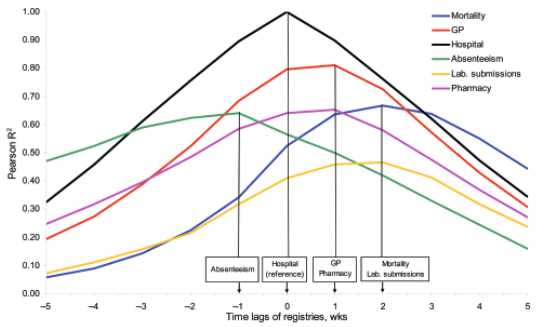


Figure 2. The (maximum) R^2 by the lagged syndromes with the hospital syndrome as a reference. Aggregated by week, univariate Pearson correlation coefficients were calculated of the hospital syndrome and each of the other syndromes. Note that the Pearson correlation coefficients are calculated over different 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

Chapter 3

Detection of Excess Influenza Severity: Associating Respiratory Hospitalization and Mortality Data With Reports of Influenza-Like Illness by Primary Care Physicians

Am J Public Health [Epub ahead of print]

Detection of Excess Influenza Severity: Associating Respiratory Hospitalization and Mortality Data With Reports of Influenza-Like Illness by Primary Care Physicians

Cees C. van den Wijngaard, MSc, Liselotte van Asten, PhD, Adam Meijer, PhD, Wilfrid van Pelt, PhD, Nico J.D. Nagelkerke, PhD, Gé A. Donker, PhD, Marianne A.B. van der Sande, PhD, and Marion P.G. Koopmans, DVM, PhD

Objectives. We explored whether excesses in influenza severity can be detected by combining respiratory syndromic hospital and mortality data with data on influenza-like illness (ILI) cases obtained from general practitioners.

Methods. To identify excesses in the severity of influenza infections in the population of the Netherlands between 1999 and 2005, we looked for increases in influenza-associated hospitalizations and mortality that were disproportionate to the number of ILI cases reported by general practitioners. We used generalized estimating equation regression models to associate syndromic hospital and mortality data with ILI surveillance data obtained from general practitioners. Virus isolation and antigenic characterization data were used to interpret the results.

Results. Disproportionate increases in hospitalizations and mortality (relative to ILI cases reported by general practitioners) were identified in 2003/04 during the A/Fujian/411/02 (H3N2) drift variant epidemic.

Conclusions. Combined surveillance of respiratory hospitalizations and mortality and ILI data obtained from general practitioners can capture increases in severe influenza-associated illness that are disproportionate to influenza incidence rates. Therefore, this novel approach should complement traditional seasonal and pandemic influenza surveillance in efforts to detect increases in influenza case fatality rates and percentages of patients hospitalized. (*Am J Public Health.* Published online ahead of print September 23, 2010: e1–e7. doi:10.2105/AJPH.2009.168245)

Syndromic surveillance is increasingly used to monitor symptoms or clinical diagnoses such as shortness of breath or pneumonia as indicators of infectious disease. The primary objective of many syndromic surveillance systems is the detection of unexpected disease increases such as those that occur as a result of bioterrorism attacks or outbreaks of emerging diseases such as severe acute respiratory syndrome (SARS). However, the signals generated by such syndromic surveillance also reflect influenza activity.[1–4]

Worldwide, influenza continues to result in serious morbidity and mortality.[5,6] The recurrence of influenza epidemics is predominantly caused by both the antigenic drift of influenza viruses and changes in the dominant virus types or subtypes. Antigenic drift occurs during the replication process of influenza viruses when mutations in surface proteins lead to declines in the level of immunity acquired through natural infection or vaccination.[7] In addition, the annual variations in dominant virus types or subtypes, such as A(H1), A(H3), and B, can lead to differences in influenza-related morbidity and mortality. For example, in recent decades levels of morbidity and mortality seem to have been lower in the influenza A(H1) and B epidemic seasons than in the A(H3) seasons.[8,9]

In the Netherlands, as in many countries, surveillance of influenza is conducted by a network of sentinel general practitioners. Influenza-like illness (ILI) consultations are reported weekly, and antigenic properties of isolated viruses are analyzed to determine their effects on annual ILI fluctuations.[10,11] Such sentinel surveillance is considered adequate for monitoring the onset and magnitude of annual influenza epidemics. However, it is not sufficient for monitoring the incidence of severe influenza infections leading to hospitalization or death.

Although the relationship between the virulence and transmission capacity of influenza viruses is still incompletely understood,[7] variations in virulence may result in disproportionate increases in severe illness relative to increases in the number of patients with ILI consulting their general practitioners. Such increases might be captured by monitoring temporal changes in the association of ILI data obtained from general practitioners (hereafter GP–ILI data) with hospitalization and mortality surveillance data. Such monitoring is not a part of current global influenza monitoring activities, although in some countries ILI data in addition to hospitalization and mortality data are included in influenza surveillance.[12,13]

We explored the potential of this monitoring strategy to detect excesses in influenza infection severity by investigating shifts in the annual association of respiratory hospitalizations and mortality with GP–ILI incidence data in the Netherlands between 1999 and 2005. In addition, we evaluated whether such shifts were associated with reported circulation of influenza virus drift variants, mismatches with vaccine strains, or changes in dominant circulating virus types or subtypes.

METHODS

We obtained hospital and mortality data from the Dutch national medical register (99% coverage of discharge and secondary diagnoses by date of hospitalization) and the Dutch causes of death registry (100% coverage of primary causes of death, as well as complicating causes and other additional causes of death, by date of death). We formulated respiratory hospitalization and mortality syndrome definitions guided by the syndrome definitions of the Centers for Disease Control and Prevention, as coded in the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*; see Appendix A, available as an online supplement to this article at http://www.ajph.org).[14] It has been demonstrated that these respiratory hospitalization and mortality syndrome data reflect respiratory pathogen activity as measured via laboratory counts.[4]

We collected ILI data from a sentinel network of general practitioners.[10] Data on influenza viruses detected in the Netherlands between 1999 and 2005 were derived from the Dutch influenza surveillance consortium (comprising the National Influenza Centre and the Netherlands Institute for Health Services Research).[11,15]

We used weekly counts of various respiratory pathogens to adjust for the effects on respiratory hospitalizations and mortality of pathogen activity other than influenza. We collected data on respiratory syncytial virus (RSV), rhinovirus, *Mycoplasma pneumoniae*, parainfluenza virus, enterovirus, and adenovirus pathogen counts from a routine laboratory surveillance system (the Weekly Sentinel Surveillance System of the Dutch Working Group on Clinical Virology, which covers 38%–73% of the population of the Netherlands).[16] We also used national mandatory notifications to obtain weekly pertussis counts.

Data Analysis

With the exception of laboratory pathogen counts (for which data on age were not available), we aggregated data by week and age category (0–4 years, 5–19 years, 20–64 years, 65 years or older). In our analyses, we excluded respiratory mortality among those in the 0- to 4-year and 5- to 19-year age groups because of the sporadic counts in these groups. SAS version 9.1 (SAS Institute Inc, Cary, NC) was used in conducting the analyses.

For the general practitioner, hospital, and mortality data, we calculated incidence rates instead of counts to quantify risk differences between age categories and to correct for changes in the age distribution of the population and (for the general practitioner sentinel data) changing registry coverage over time. After plotting time series of GP–ILI, respiratory hospitalization, and mortality incidence data, we looked for increases in hospitalizations and mortality that seemed disproportionate to increases in ILI cases as a measure of severity of illness. We also examined time series of respiratory pathogens other than influenza to assess whether elevations in respiratory hospitalizations or mortality might be associated with other pathogen activity (measured via routine laboratory surveillance).

We used additive generalized estimating equation (GEE)[17] models with a Poisson error distribution to detect elevations in respiratory hospitalizations and mortality that were disproportionate to seasonal rises in ILI incidence in the general practitioner sentinel data. We estimated hospitalization and mortality time series, stratified by age, according to lagged ILI incidence. We then used the 95% upper limits of the models (details on the model variables are provided in Appendix B, available as an online supplement to this article at http://www.ajph.org) to determine distinct episodes in time in which hospitalizations and mortality increases were disproportionate to average modeled associations with ILI incidence rates.

To adjust for the activity of respiratory pathogens other than influenza, we considered RSV, rhinovirus, parainfluenza virus, *M pneumoniae*, adenovirus enterovirus, and pertussis counts for inclusion in the models as well. We also assumed a constant basic syndrome level attributable to factors other than respiratory pathogen activity. In the summer months, however, as the basic syndrome level appeared to be lower with respect to hospitalizations (possibly as a result of fewer planned hospitalizations during that period), we used a lower basic syndrome level (by including a dummy variable for "summer"). The regression model coefficients for each of the lagged pathogens and for the GP–ILI incidence data were assumed to be constant in time.

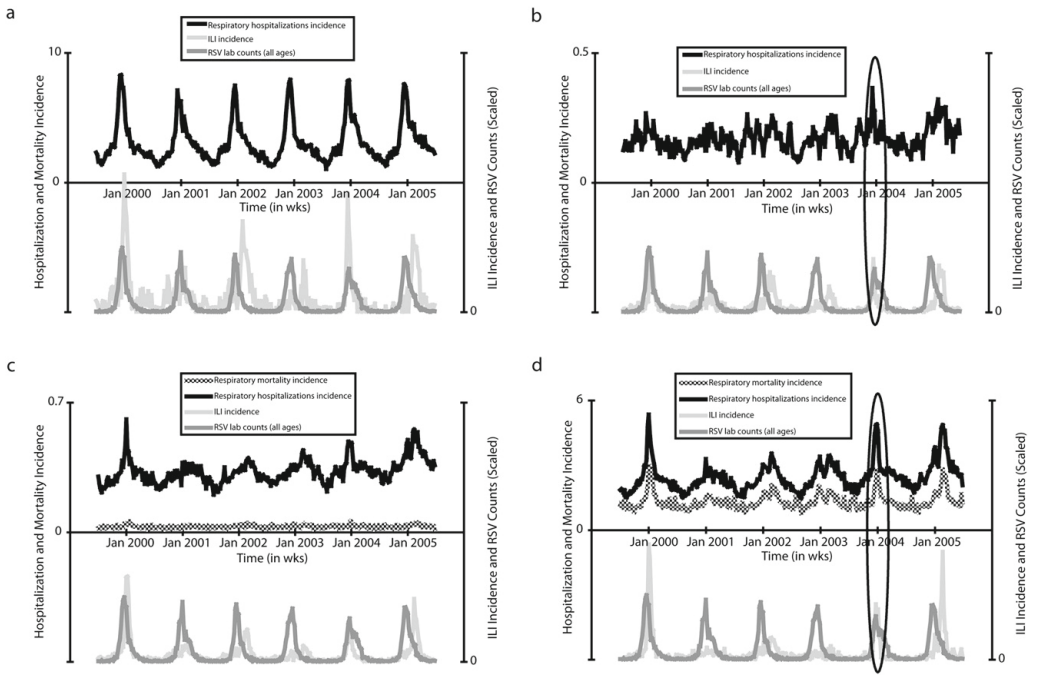
We initially built a generalized linear model with a Poisson error distribution and an identity link. To do so we used a forward stepwise regression approach, selecting the lagged ILI incidence and lagged pathogen counts that contributed most to the model fit (5-week lags were used; e.g., in step 1, ILI was included with a 1-week lag if that exhibited a better model fit than all other pathogen–lag combinations, assessed with Akaike's information criterion[18]). We included lagged GP–ILI incidence and the counts for each pathogen in the model only once and only if results were significant at the $P \leq .05$ level.

We analyzed age-stratified hospitalization, mortality, and GP–ILI incidence data in the regression models. Age stratification was not possible for pathogen counts. We excluded negative associations for pathogen counts to avoid spurious model fits due to biologically implausible associations (e.g., negative associations between enterovirus, which peaks in summer, and respiratory syndromes, which peak in winter). Also, we added seasonal variables (sine and cosine terms)—guided by periodograms of the model residuals, which reflect the importance of specific cyclical periods (e.g., 26 weeks, 52 weeks) in explaining the variance in the residuals—to correct for seasonal variation and used GEEs to correct the model outcomes for autocorrelation between observations.

To quantify temporal heterogeneity in associations of GP–ILI data with data on hospitalizations and mortality respiratory syndromes, we modified the models by using time-dependent (by epidemic year, defined as July 1 through June 30) ILI regression coefficients (instead of a single regression coefficient for all years). These annual ILI regression coefficients can be seen as scaling factors for the number of hospitalizations or deaths associated with a one-case increase in ILI incidence per 10 000 population. We plotted estimates for these coefficients on a bar chart. The years with the highest estimated ILI regression coefficients were considered as those associated with the most severe illness per ILI case. We conducted F tests (with a null hypothesis of no differences in associations over the study period) to determine whether the coefficients differed across the years of the study (Appendix B, available as an online supplement to this article at http://www.ajph.org).

Influenza Virus Isolation and Antigenic Characterization

To assess whether disproportionate levels of respiratory hospitalizations and mortality (relative to GP–ILI incidence rates) might be related to the circulation of specific influenza virus variants or subtypes, we explored weekly reports of influenza virus subtypes A(H1), A(H3), and B and assessed, on the basis of antigenic characterization, which influenza virus drift strains were present in the Netherlands during 1999–2005.



Note. RSV = respiratory syncytial virus; ILI = influenza-like illness. No mortality time series were plotted for the 0–4-year and 5–19-year age groups because of their low numbers. RSV counts were plotted for all age groups (because no age data were available), and the counts were scaled to fit the graph.

FIGURE 1—Respiratory hospitalizations and mortality incidences versus ILI incidence rates and RSV laboratory counts, by age group (a) 0–4 years, (b) 5–19 years, (c) 20–64 years, and (d) 65 years or older: the Netherlands, 1999–2005.

We also evaluated to what extent these drift strains were reported to match or not match the vaccine strains for those years. Individuals at increased risk for complications of influenza (elderly people and those with specific comorbid conditions) are offered annual influenza vaccination[19] (during the study period, vaccine coverage levels in the Netherlands were above 65% for individuals aged 65 years or older and approximately 80% for those aged older than 75 years). Data on antigenic characteristics and the match between vaccines and circulating viruses were derived from annual influenza surveillance reports.[11] To assess other possible explanations for disproportionate levels of respiratory hospitalizations and mortality relative to GP–ILI incidence rates, we also compared the analysis results against plotted time series of specific morbidity patterns associated with respiratory hospitalizations, as measured via *ICD-9*-coded hospital diagnosis incidence rates.

RESULTS

Plots of GP–ILI time series and respiratory hospitalization and mortality time series showed approximately concurrent peaks in all winter seasons. The highest peaks were observed in 1999/00 and 2004/05 (data not shown). The influenza epidemics in 2000/01, 2001/02, and 2002/03 were relatively mild. When data on respiratory pathogen counts (other than influenza) were plotted (data not shown), RSV showed the clearest winter peaks, concurrent with elevations in respiratory hospitalizations and mortality.

Therefore, we plotted respiratory hospitalizations and mortality against GP–ILI incidence rates and laboratory RSV counts stratified by age (Figure 1). Elevations in respiratory hospitalizations were highest in the youngest and oldest age groups (0–4 years and 65 years or above), and elevations in respiratory mortality were highest in the oldest age group. Hospitalizations in the 0–4-year age group corresponded more with the RSV time series than with the ILI time series (Figure 1). During the 2003/04 winter season, steep peaks in respiratory hospitalizations were observed among those aged 5 to 19 years and those 65 years or older, and (although RSV counts peaked at the same time) these peaks seemed disproportionately high relative to the ILI time series during that season (Figure 1). This trend was also observed for mortality among individuals 65 years or older (Figure 1, indicated by ellipse).

Regression Analysis

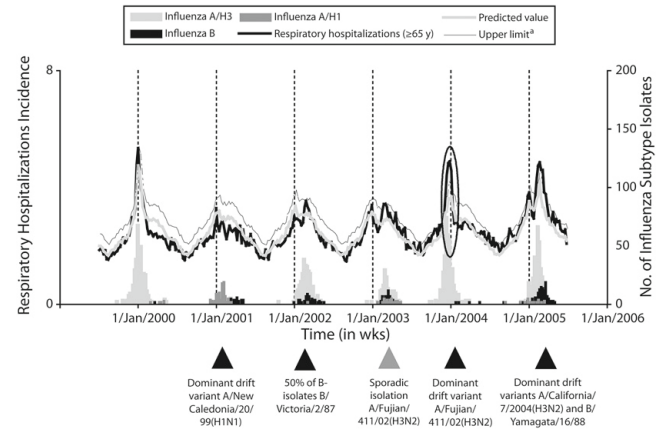
In the models with constant ILI regression coefficients over the entire study period, variations in respiratory hospitalizations and mortality among individuals 65 years or older and variations in hospitalizations among those aged 0 to 4 years were explained quite well by variations in ILI incidence and respiratory pathogen counts. The explained variance was lower for the other age groups.

Periodograms of the model residuals showed sharp peaks at 1 year along with smaller harmonics for shorter periods. We therefore added sine and cosine terms to the models to adjust for seasonal trends, and we used GEEs to correct for autocorrelation in the residuals. These model refinements led to only minimal changes in the ILI regression coefficients and the explained variance of the models; percentages of explained variance for hospitalizations were 95% among those aged 0 to 4 years, 47% among those aged 5 to 19 years, 68% among those aged 20 to 64 years, and 78% among those 65 years or older. Percentages of explained variance for mortality were 37% among those aged 20 to 64 years and 76% among those 65 years or older.

With respect to periods of peak influenza activity (as measured by peaks in GP–ILI incidence concurrent with peaks in the counts of influenza isolates), the time series of actual hospitalizations among both those aged 0 to 4 years (data not shown) and those 65 years or older (Figure 2) most clearly exceeded the 95% upper limit of the models during winter 2003/04. A subsequent (F-test) analysis of the model in which year-specific ILI regression coefficients were used showed significant annual heterogeneity in these coefficients for all age categories ($P \leq .001$).

Figure 3a shows the annual GP–ILI regression coefficients for respiratory hospitalizations. For example, the regression coefficient value of 3.94 for hospitalizations in the 0- to 4-year age group in 2003/04 indicates that, for a hypothetical ILI incidence of 100 per 10 000 population, the estimated respiratory hospitalization incidence for that age group is 3.94 (per 10 000 population). The annual GP–ILI regression coefficients for respiratory hospitalizations were highest among those 65 years or older and those aged 0 to 4 years. In addition, the regression coefficients for these age groups were significantly higher in 2003/04 than in any other study year ($P \leq .001$).

Figure 3 (panel b) shows that, as expected, the mortality regression coefficient was much higher for those 65 years or older than for those aged 20 to 64 years. Similar to the data for hospitalizations, the ILI regression coefficient for those 65 years or older was clearly higher in 2003/04 than in any other study year ($P \leq .03$). In 2000/01, some of the estimated ILI regression coefficients were below zero, reflecting the mild influenza impact in that season.



Note. The hospitalization incidence for individuals aged 65 years or older is plotted in a line graph with the predicted value and the 95% upper limit (Appendix B, available as an online supplement at http://www.ajph.org). Values exceeding the model's upper limit are indicated by the ellipse. Below the line graphs, the counts of influenza isolates by subtype—A(H3), A(H1), and B—are presented as bars on the x-axis, and reports of drift variants are indicated.
[a]Because we used generalized estimating equation models, confidence intervals for prediction were not available.

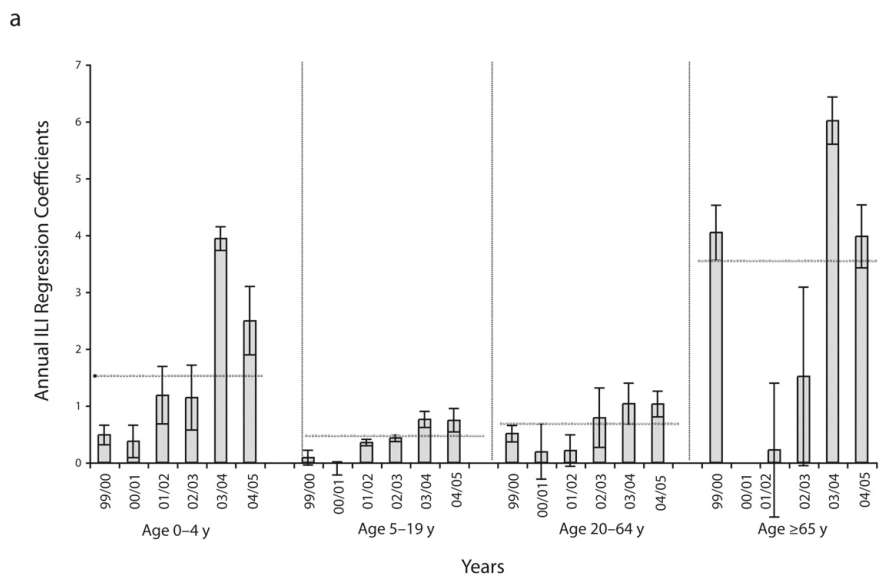
FIGURE 2—Respiratory hospitalization incidence explained by influenza-like illness incidence versus influenza virus subtype counts and reports of drift variants: the Netherlands, 1999–2005.

Influenza Virus Isolation and Antigenic Characterization

Figure 2 presents data on influenza virus subtypes and reported introductions of drift variants.[11,20–22] All reported influenza drift strains mismatched to some extent with the vaccine strains observed over the study period with the exception of the Caledonia/20/99 (H1N1) strain in 2000/01.

Specific Hospital Diagnoses

In visually exploring respiratory hospitalization discharges and diagnoses (data not shown), we focused on elevations in time that may have been related to the excess number of respiratory hospitalizations observed in 2003/04. The elevations in hospitalizations involving a diagnosis of pneumococcal pneumonia (*ICD-9* code 481) or pneumonia due to streptococcus (*ICD-9* code 4823) during peak winter influenza activity in 2003/04 and, to a lesser extent, 2002/03 were among the highest observed in the study period (in 2002/03, as a percentage of respiratory hospitalizations overall, pneumococcal pneumonia showed the highest elevation over the study period). The second highest elevation in hospitalizations involving an influenza diagnosis (*ICD-9* codes 4870 and 4871) was observed in 2003/04 (the highest elevation was in 1999/00; no significant elevations were observed in 2002/03).



Note. The horizontal line in each chart gives the value of an ILI regression coefficient that is constant in time, as an indication of the average value of the ILI regression coefficients over the study period. The 95% confidence intervals for the regression coefficients are presented in the figure as well.

FIGURE 3—Annual (July 1–June 30) estimates of the association of influenza-like illness (ILI) incidence with (a) respiratory hospitalization incidence in all age groups and (b) respiratory mortality incidence for individuals aged 20–64 years and individuals 65 years or older: the Netherlands, 1999–2005.

DISCUSSION

We observed increases in severe illness due to influenza in the Netherlands between 1999 and 2005 that were disproportionate to ILI incidence rates. Our observations reveal the existence of temporal heterogeneities in the severity of influenza infections, possibly stemming from variations in the virulence of circulating influenza viruses. Several studies have shown that syndromic data on general respiratory symptoms and clinical diagnoses can be useful in influenza surveillance.[2,3,23–25] We combined respiratory syndrome data on hospitalizations and mortality with traditional ILI surveillance data obtained from general practitioners to determine year-to-year differences in the number of respiratory hospitalizations and deaths in proportion to the number of ILI cases.

We linked our observations to virological changes by visually exploring time series of influenza subtype counts and reported antigenic information about influenza virus strains. Five drift variants were reported in the period under study—2 A(H3) variants, 1 A(H1) variant, and 2 B variants (Figure 2)—but only in 2003/04, in the case of A/Fujian/411/02(H3N2),[20] did this reporting of a drift variant concur with disproportionate levels of hospitalizations and mortality.

Although at first glance these results seem to suggest that it is difficult to predict clinical effects from virological data, a more thorough look at our virological findings explains the absence of excess effects in years other than 2003/04. The relatively low hospitalization and mortality levels in comparison with ILI incidence rates in 2000/01 and 2001/02 can be explained by the relative lack of fitness of the A/New Caledonia/20/99(H1N1) and B/Victoria/2/87 variants (respectively), as morbidity and mortality levels tend to be lower in seasons with predominantly A(H1)[21] or B[22] strains than in A(H3) seasons.[8,9]

In addition, the A/New Caledonia/20/99(H1N1) drift variant reported during 2000/01 had emerged in 1999/00, and the vaccine for the 2000/01 season contained this strain and probably provided optimal protection against this drift variant, thereby reducing severe illness in elderly people who had been vaccinated.[21] In 2004/05, influenza A(H3) and influenza B drift strains were reported,[11] but their impact was only moderate. During this season, the antigenic distance of the dominant A/California/7/04-like(H3N2) drift variant virus toward the influenza A(H3N2) virus in 2003/04 (A/Fujian/411/02-like) was relatively small.[11]

This was not the case for A/Fujian/411/02-like(H3N2) viruses in 2003/04, which were quite distinct from preceding A(H3N2) viruses,[11] thereby representing a likely explanation for the observed excess hospitalizations and mortality in that flu season. The high influenza impact among young children and the elderly, relative to the limited size of the 2003/04 epidemic measured according to GP-ILI data, seems to be consistent with the high hospitalization rates during the 2003 influenza season in New Zealand in combination with the limited size of the epidemic also according to GP-ILI data.[26] There A/Fujian/411/02(H3N2) was the dominant subtype as well.

Some other European countries reported dominant activity or more severe outbreaks of A/Fujian/411/02(H3N2) in 2002/03, but there was great variation across Europe in circulating strains during that winter.[27,28] In the Netherlands, A/Fujian/411/02(H3N2) strains were also circulating in that period, but they were isolated only sporadically; 5 isolates were observed, accounting for 4% of A(H3N2) isolates overall.[29]

The introduction of new influenza drift variants and shifts in influenza subtypes are not the only possible explanations for the observed differences in influenza impact. Other viral factors (e.g., viral replication capacity, virulence, viral transmissibility) and climatic factors (e.g., temperature and relative humidity) may likewise influence the impact of seasonal influenza on morbidity and mortality. For instance, studies have suggested that the antigenic drift of the A(H3N2) viruses reported in 2003/04 resulted in declines in the level of population immunity (leading to A/Fujian/411/02 in 2002) but that this drift variant became widespread only after gaining a higher viral replication capacity through additional reassortment-related changes in internal genes.[30–32]

Limitations

A limitation of this study is that it was based on associating time series of hospitalizations and mortality with ILI data, and such associations could be confounded by seasonally circulating pathogens other than influenza. To minimize this possibility, we adjusted for the possible impact of RSV and other respiratory pathogens by including them in our regression models. We also included seasonal terms to correct for possible confounding by other seasonally varying factors. Our use of autocorrelation in our models corrected for other, possibly transient causes of hospitalization and mortality.

Another observation lent additional support for the association of influenza with excess elevations. That is, in 2003/04, concurrent with a moderately high ILI peak, hospitalizations involving an influenza diagnosis exhibited the second highest elevation over the study, and hospital diagnoses of pneumococcal pneumonia showed a high elevation as well (which seems to be in line with observations that influenza infections may predispose patients for *S pneumoniae* infections[33–35]).

Also, to enhance prospective surveillance, there is a need to further evaluate how increases in hospitalizations and mortality that are disproportionate to ILI incidence rates can be detected on a timely basis within a particular influenza season. Quality control chart approaches[36,37] might be developed for the timely detection of such temporal changes that require the attention of health authorities.

Conclusions

Our results show that increases in severe influenza-associated illness that are disproportionate to the incidence of influenza in the community can be detected through combined analyses of GP–ILI data and data on respiratory hospitalizations and mortality. This novel approach should be implemented in global influenza surveillance programs to provide better estimates of increases in severe morbidity and mortality due to influenza infections.

Our data also show that there is a possible relationship between influenza impact and specific influenza strains. Further research is needed to better understand the causes of such relationships. It seems worthwhile to develop prospective respiratory syndromic surveillance of hospitalizations and mortality complementing traditional seasonal and pandemic influenza surveillance to allow detection of increases in influenza case fatality rates and percentages of patients hospitalized. During ongoing (pandemic) influenza epidemics, such surveillance

information could be used to determine the need for control measures such as additional vaccination or prophylactic treatment. ■

About the Authors

Cees C. van den Wijngaard, Liselotte van Asten, Adam Meijer, Wilfrid van Pelt, Marianne A.B. van der Sande, and Marion P.G. Koopmans are with the National Institute for Public Health and the Environment, Centre for Infectious Disease Control, Bilthoven, the Netherlands. Marion P.G. Koopmans is also with the Erasmus Medical Centre, Rotterdam, the Netherlands. Nico J.D. Nagelkerke is with the Department of Community Medicine, United Arab Emirates University, Al-Ain. Gé A. Donker is with the Netherlands Institute for Health Services Research, Utrecht.

Correspondence should be sent to Cees C. van den Wijngaard, MSc, Centre for Infectious Disease Control, Netherlands National Institute for Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, the Netherlands (e-mail: kees.van.den.wijngaard@rivm.nl). Reprints can be ordered at http://www.ajph.org by clicking on the "Reprints/Eprints" link.

This article was accepted November 4, 2009.

Contributors

C.C. van den Wijngaard contributed to the study design, analyzed the data, and wrote the article. L. van Asten contributed to the study design, data analysis, and drafting of the article. A. Meijer interpreted the virological data and helped draft the article. W. van Pelt and M.P.G. Koopmans contributed to the study design and the interpretation of data and reviewed drafts of the article. N.J.D. Nagelkerke contributed to the statistical design and the interpretation of data and reviewed drafts of the article. G. Donker collected surveillance data and reviewed drafts of the article. M.A.B. van der Sande helped interpret the results and reviewed drafts of the article.

Acknowledgments

We thank Statistics Netherlands, the Dutch National Medical Register, the Netherlands Institute for Health Services Research, and the National Influenza Centre of the Netherlands for providing data; the members of the Dutch Working Group on Clinical Virology for allowing us access to their registry and for collecting and providing weekly diagnostic results; Hans van Vliet and Mirjam Kretzschmar for reading and commenting on the article; and Hendriek Boshuizen and Jan van de Kassteele for providing feedback on the statistical methods used.

Human Participant Protection

Because the data examined in this study were obtained from surveillance or medical research registries, no protocol approval was needed.

References

1. Buehler JW, Berkelman RL, Hartley DM, et al. Syndromic surveillance and bioterrorism-related epidemics. *Emerg Infect Dis.* 2003;9(10):1197–1204.

2. Lazarus R, Kleinman KP, Dashevsky I, et al. 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. Heffernan R, Mostashari F, Das D, et al. Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis.* 2004;10(5):858–864.

4. Van den Wijngaard C, van Asten L, van Pelt W, et al. Validation of syndromic surveillance for respiratory pathogen activity. *Emerg Infect Dis.* 2008;14(6):917–925.

5. Nicholson KG, Wood JM, Zambon M. Influenza. *Lancet.* 2003;362(9397):1733–1745.

6. Palese P. Influenza: old and new threats. *Nat Med.* 2004;10(suppl 12):S82–S87.

7. Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. *Vaccine.* 2007;25(39–40):6852–6862.

8. Simonsen L, Clarke MJ, Williamson GD, et al. The impact of influenza epidemics on mortality: introducing a severity index. *Am J Public Health.* 1997;87(12):1944–1950.

9. Chowell G, Miller MA, Viboud C. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiol Infect.* 2008;136(6):852–864.

10. Donker GA. Continuous morbidity registration sentinels: Netherlands 2006 [in Dutch]. Available at http://www.nivel.nl/pdf/CMR-Peilstations-2006.pdf. Accessed August 5, 2010.

11. De Jong JC, Rimmelzwaan GF, Bartelds AI, et al. The influenza season 2004/'05 in the Netherlands with the largest epidemic of the last 5 years caused by the virus variant A/California and the composition of the vaccine for the season 2005/'06 [in Dutch]. *Ned Tijdschr Geneeskd.* 2005;149(42):2355–2361.

12. Thompson WW, Comanor L, Shay DK. Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease. *J Infect Dis.* 2006;194(suppl 2):S82–S91.

13. Huang QS, Lopez L, Adlam B. Influenza surveillance in New Zealand in 2005. *N Z Med J.* 2007;120(1256):U2581.

14. *International Classification of Diseases, Ninth Revision, Clinical Modification.* Hyattsville, MD: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.

15. Dijkstra F, van der Plas SM, Meijer A, et al. Respiratory surveillance in 2004/2005 [in Dutch]. Available at: http://www.rivm.nl/infectieziektenbulletin/bul1606/art_surveillance.html. Accessed August 5, 2010.

16. Van den Brandhof WE, Kroes ACM, Bosman A, et al. Reporting virus diagnostics in the Netherlands: representativeness of the virological weekly reports [in Dutch]. Available at: http://www.rivm.nl/infectieziektenbulletin/bul1304/vir_diagnostiek.html. Accessed August 5, 2010.

17. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73(1):13–22.

18. Akaike H. A new look at statistical model identification. *IEEE Trans Automat Contr.* 1974;19(6):716–723.

19. Hak E, van Loon S, Buskens E, et al. Design of the Dutch Prevention of Influenza, Surveillance and Management (PRISMA) study. *Vaccine.* 2003;21(15):1719–1724.

20. Rimmelzwaan GF, de Jong JC, Bartelds AI, et al. The 2003/2004 influenza season in the Netherlands with a limited epidemic of the virus variant A/Fujian, and the vaccine composition for the 2004/2005 season [in Dutch]. *Ned Tijdschr Geneeskd.* 2004;148(40):1984–1988.

21. De Jong JC, Rimmelzwaan GF, Bartelds AI, et al. 2000/01 influenza season and the vaccine composition for the season 2001/'02 [in Dutch]. *Ned Tijdschr Geneeskd.* 2001;145(40):1945–1950.

22. Rimmelzwaan GF, de Jong JC, Bartelds AI, et al. The 2001/2002 influenza season and the vaccine composition for the 2002/2003 season [in Dutch]. *Ned Tijdschr Geneeskd.* 2002;146(39):1846–1850.

23. Quenel P, Dab W, Hannoun C, et al. Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics. *Int J Epidemiol.* 1994;23(4):849–855.

24. Cooper DL, Smith GE, Hollyoak VA, et al. Use of NHS Direct calls for surveillance of influenza—a second year's experience. *Commun Dis Public Health.* 2002;5(2):127–131.

25. Olson DR, Heffernan RT, Paladini M, et al. Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. *PLoS Med.* 2007;4(8):e247.

26. Lopez L, Huang S, Baker M. Influenza in New Zealand—2003. Available at: http://www.surv.esr.cri.nz/PDF_surveillance/Virology/FluAnnRpt/InfluenzaAnn2003.pdf. Accessed August 5, 2010.

27. Paget WJ, Meerhoff TJ, Rebelo de Andrade H. Heterogeneous influenza activity across Europe during the winter of 2002–2003. *Euro Surveill.* 2003;8(12):230–239.

28. Bragstad K, Jorgensen PH, Handberg KJ, et al. New avian influenza A virus subtype combination H5N7 identified in Danish mallard ducks. *Virus Res.* 2005;109(2):181–190.

29. De Jong JC, Rimmelzwaan GF, Bartelds AI, et al. The 2002/2003 influenza season in the Netherlands and the vaccine composition for the 2003/2004 season. [in Dutch] *Ned Tijdschr Geneeskd.* 2003;147(40):1971–1975.

30. Holmes EC, Ghedin E, Miller N, et al. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* 2005;3(9):e300.

31. Barr IG, Komadina N, Hurt AC, et al. An influenza A(H3) reassortant was epidemic in Australia and New Zealand in 2003. *J Med Virol.* 2005;76(3):391–397.

32. Jin H, Zhou H, Liu H, et al. Two residues in the hemagglutinin of A/Fujian/411/02-like influenza viruses are responsible for antigenic drift from A/Panama/2007/99. *Virology.* 2005;336(1):113–119.

33. Kim PE, Musher DM, Glezen WP, et al. Association of invasive pneumococcal disease with season, atmospheric conditions, air pollution, and the isolation of respiratory viruses. *Clin Infect Dis.* 1996;22(1):100–106.

34. Avadhanula V, Rodriguez CA, Devincenzo JP, 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(4):1629–1636.

35. Jansen AG, Sanders EA, van der Ende A, et al. Invasive pneumococcal and meningococcal disease: association with influenza virus and respiratory syncytial virus activity? *Epidemiol Infect.* 2008;136(11):1448–1454.

36. Jarrett JE, Pan X. The quality control chart for monitoring multivariate autocorrelated processes. *Comput Stat Data Anal.* 2007;51(8):3862–3870.

37. Testik MC. Model inadequacy and residuals control charts for autocorrelated processes. *Qual Reliab Eng Int.* 2005;21(2):115–130.

Chapter **4**

Syndromic Surveillance for Local Outbreaks of Lower-Respiratory Infections: Would It Work?

PLoS ONE 2010;5:e10406

Syndromic Surveillance for Local Outbreaks of Lower-Respiratory Infections: Would It Work?

Cees C. van den Wijngaard[1]*, Liselotte van Asten[1], Wilfrid van Pelt[1], Gerda Doornbos[1], Nico J. D. Nagelkerke[2], Gé A. Donker[3], Wim van der Hoek[1], Marion P. G. Koopmans[1,4]

1 Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands, **2** United Arab Emirates University, Al-Ain, United Arab Emirates, **3** Netherlands Institute of Health Services Research, Utrecht, The Netherlands, **4** Erasmus Medical Center, Rotterdam, The Netherlands

Abstract

Background: Although syndromic surveillance is increasingly used to detect unusual illness, there is a debate whether it is useful for detecting local outbreaks. We evaluated whether syndromic surveillance detects local outbreaks of lower-respiratory infections (LRIs) without swamping true signals by false alarms.

Methods and Findings: Using retrospective hospitalization data, we simulated prospective surveillance for LRI-elevations. Between 1999–2006, a total of 290762 LRIs were included by date of hospitalization and patients place of residence (>80% coverage, 16 million population). Two large outbreaks of Legionnaires disease in the Netherlands were used as positive controls to test whether these outbreaks could have been detected as local LRI elevations. We used a space-time permutation scan statistic to detect LRI clusters. We evaluated how many LRI-clusters were detected in 1999–2006 and assessed likely causes for the cluster-signals by looking for significantly higher proportions of specific hospital discharge diagnoses (e.g. Legionnaires disease) and overlap with regional influenza elevations. We also evaluated whether the number of space-time signals can be reduced by restricting the scan statistic in space or time. In 1999–2006 the scan-statistic detected 35 local LRI clusters, representing on average 5 clusters per year. The known Legionnaires' disease outbreaks in 1999 and 2006 were detected as LRI-clusters, since cluster-signals were generated with an increased proportion of Legionnaires disease patients (p:<0.0001). 21 other clusters coincided with local influenza and/or respiratory syncytial virus activity, and 1 cluster appeared to be a data artifact. For 11 clusters no likely cause was defined, some possibly representing as yet undetected LRI-outbreaks. With restrictions on time and spatial windows the scan statistic still detected the Legionnaires' disease outbreaks, without loss of timeliness and with less signals generated in time (up to 42% decline).

Conclusions: To our knowledge this is the first study that systematically evaluates the performance of space-time syndromic surveillance with nationwide high coverage data over a longer period. The results show that syndromic surveillance can detect local LRI-outbreaks in a timely manner, independent of laboratory-based outbreak detection. Furthermore, since comparatively few new clusters per year were observed that would prompt investigation, syndromic hospital-surveillance could be a valuable tool for detection of local LRI-outbreaks.

Citation: van den Wijngaard CC, van Asten L, van Pelt W, Doornbos G, Nagelkerke NJD, et al. (2010) Syndromic Surveillance for Local Outbreaks of Lower-Respiratory Infections: Would It Work? PLoS ONE 5(4): e10406. doi:10.1371/journal.pone.0010406

Editor: Rory Edward Morty, University of Giessen Lung Center, Germany

Received January 6, 2010; **Accepted** April 6, 2010; **Published** April 29, 2010

Funding: This study was funded by the Dutch Ministry of Health, Welfare and Sport. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kees.van.den.wijngaard@rivm.nl

Introduction

The SARS epidemic in 2003, the bioterrorism attacks in 2001, and the ongoing threat of new infectious disease outbreaks have prompted many countries to invest in their capacity to respond timely to emerging infectious disease outbreaks, as early outbreak-detection may well mitigate their impact. As a result, new surveillance systems for earlier detection have been implemented, often labeled "syndromic surveillance" [1–6]. These systems use increased reporting of critical symptoms or clinical diagnoses as early indicators of infectious disease outbreaks. This not only allows monitoring of clinical syndromes before laboratory diagnoses have been made, but also allows detection of outbreaks of diseases for which no diagnostics were requested or available (including emerging pathogens). Geographic analysis methods – such as space-time scan statistics – may further increase the sensitivity of syndromic surveillance for detection of local outbreaks or of regional differences in regular seasonal epidemic diseases [2,6]. In the SARS outbreak in Hongkong in 2003, it is believed that a near real-time space-time analysis would have detected the highly unusual clustering of severe acute respiratory syndrome cases much sooner [7]. However, concerns exist about the specificity of space-time syndromic surveillance, i.e. that it might generate many false signals [8,9].

The objective of this study was to evaluate to what extent syndromic surveillance detects local outbreaks of lower-respiratory infections (LRIs) without swamping true signals by false alarms. Using retrospective hospitalization data, we simulated prospective space-time syndromic surveillance for LRI-elevations. The two largest outbreaks of Legionnaires' disease in the Netherlands in the last decade were used as "positive controls" to test whether these known outbreaks would have been detected by space-time signals in LRI data. To assess other (likely) causes for detected LRI-elevations, we examined regional increases in the reported incidence of influenza-like-illness (ILI), hospital discharge diagnoses for respiratory illnesses and age group distributions for LRI cases. We also evaluated whether the number of generated space-time signals can be reduced by restricting the time and spatial windows for the analyses.

Methods

Ethical Approval

Since we only used anonymous data from existing medical research and surveillance registries, neither formal ethics committee approval nor informed consent from the patients were required.

LRI-syndrome data (1999–2006)

Hospitalization data were collected from the Dutch National Medical Register (discharge and secondary diagnoses by date of hospitalization for 1999–2006). In 1999–2004 this registry had a 99% coverage (16 million pop.) and in 2005/6 approximately 80%, after exclusion of hospitals with incomplete data for those years.

We included all records on hospitalizations with any kind of LRI as either discharge or secondary diagnosis, under the assumption that this reflects prospective classification of patients with a lower respiratory infection in a "LRI-syndrome" on the day of hospitalization. ICD-9-CM (International Classification of Diseases, 9th revision, Clinical Modification) codes for a LRI syndrome were selected from the CDC respiratory syndrome codes-list (Centers for Disease Control and Prevention, USA, http://www.bt.cdc.gov/surveillance/syndromedef/; and see Appendix S1). After excluding duplicate hospitalizations of the same patient within 6 weeks (5% excluded), 222638 records were included for 1999–2004, and 68124 for 2005–2006. Data were aggregated by hospitalization date, postal-code and age group (0–4, 5–19, 20–49, 50–64, ≥65 years). Since higher levels of spatial resolution can result in more sensitive detection of outbreaks [10,11] we used 4-digit postal-codes (4023 areas in a 16 million population), which provide the highest level of spatial resolution available within privacy regulations.

Regional ILI-surveillance data

ILI-data were collected from a sentinel network of general practitioners (GPs, Continuous Morbidity Registration Centres, CMR sentinel stations, 1% population coverage) [12]). The ILI-counts and underlying GP-practice populations were aggregated by region and week. The GP-practice populations were corrected for weeks that specific GP-practices did not supply data. Due to the small number of GP-practices in some parts of the country, the weekly ILI-data were aggregated in 4 major regional groups instead of postal codes.

Test-case outbreaks

Two large outbreaks of Legionnaires' disease were used as "positive controls" for emerging LRI-outbreaks [13,14]:

1) In March 1999, a large Legionnaires' disease outbreak occurred among persons who had visited a flower show [13]. Ten patients with pneumonia were admitted to one hospital between March 7th to 11th. By March 11th, six patients were diagnosed with Legionnaires' disease and an alarm notice was given to hospitals and GPs in the region. Follow-up investigation detected a total of 188 cases, of whom 167 (87%) were hospitalized and 21 (11%) died.
2) Between July 6th–28th 2006, 30 Legionnaires' disease cases were identified in Amsterdam, 2 of which were fatal [14]. On July 7th an alarm notice was given. A cooling tower in the town centre was later identified as the outbreak-source.

Scan statistics for space-time clusters

For the LRI-data, we used a space-time permutation scan statistic which compared the observed number of cases in circular areas with variable radii in flexible time periods vs the expected number of cases, based on the geographic distribution of cases in the whole dataset [15]. In this way, only the case data is needed to estimate the expected number of cases in each space-time window, and population density and time trends in the case data are automatically adjusted for.

We used SaTScan software [16] and the SaTScan Macro Accessory for Cartography (SMAC [17], applied in SAS version 9.1, SAS Institute Inc., Cary, NC, USA) to run the scan-statistic and visualize the results. We simulated a prospective surveillance by running the scan-statistic on data from the year preceding each time unit (day or week) in the analysis period. Thus, weekly or daily space-time signals were generated, each time that the observed number of cases in a certain space and time window exceeded the defined significance threshold. Since such analysis consumes a lot of computation time, we performed weekly analysis (instead of daily) over the whole study period. Daily analyses were also performed in the years that the test-case outbreaks occurred (1999 and 2006), to assess the earliest possible detection date. For all analyses, we chose to use a time-aggregation level of 7-days length. For the daily analyses, these 7-day aggregation windows shifted one day forward for each daily run. Thus we both reduced the computation time and adjusted for day-of-week effects (both purely temporal and spatial day-of-week effects).

To indicate the significance of detected space-time signals, we used recurrence intervals, which indicate how often a signal of the observed significance would be observed by chance under the hypothesis of no outbreak [18]. I.e. if the recurrence interval of a signal is say 1 year, 1 signal of the observed significance is expected in 1 year. Two thresholds levels were used: signals with recurrence interval ≥1 and ≥5 years. We assessed whether successive signals overlapped in space and time, which suggests the same cause. For the sake of readability, we indicated a group of such overlapping space-time signals as "cluster" and an individual space-time signal as "cluster-signal".

We evaluated how many LRI-clusters and signals were detected over the whole study period (1999–2006) and looked for explanations guided by the two-step criteria in Figure 1. In step one, we assessed likely causes for the cluster-signals by looking for significantly higher proportions of specific hospital discharge diagnoses (e.g., Legionnaires' disease [19,20]). In step two we assessed overlap with regional ILI clusters (Appendix S2), as (local) influenza activity might be reflected in local LRI-elevations. Since other pathogens than influenza might cause some ILI fluctuations, influenza activity was only considered to be a likely cause if space-time overlap between LRI and ILI-clusters coincided with the annual influenza season (Figure 1). If a specific cause was defined for one or more signals within one cluster, we considered that to be a likely cause for the whole cluster. We also evaluated the timeliness of detection for the clusters related to the known Legionnaires' disease outbreaks.

A sensitivity analysis was used to evaluate the impact of time and spatial window settings on the number of clusters and signals

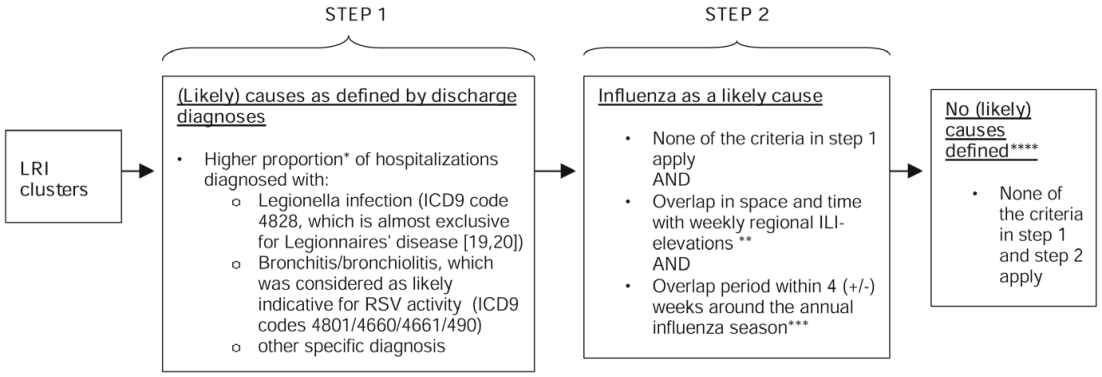


Figure 1. Two-step criteria to define (likely) causes for LRI hospitalization clusters detected in 1999–2006. * As evaluated by the right-sided Fisher's exact test for 2×2 Tables (alpha≤0.01) of hospitalizations within vs hospitalizations outside of the cluster-signal. The proportion of hospitalizations with a specific characteristic (e.g. legionnaires' disease as discharge diagnoses, or age 20–49 yrs) can be significantly higher among hospitalizations within the cluster-signal than the proportion outside of the cluster-signal. ** For the ILI-cluster-signals we could only use 4 major regions as spatial resolution. Overlap in time between LRI and ILI-cluster-signals was defined as occurrence of weekly ILI-cluster-signals within 2 weeks (+/−) around LRI-cluster-signals. ***The annual influenza season was defined as all weeks with a national weekly ILI-incidence ≥3 per 10.000 pop. **** Possibly unreported/undetected local LRI-outbreaks by undetected pathogens.
doi:10.1371/journal.pone.0010406.g001

detected. For the initial analyses, we put only minor constraints on the maximum temporal and spatial windows of the scan-statistic, to avoid wrongful assumptions about time, geographical location and size of an outbreak. We then repeated these weekly analyses with a temporal window of maximum 7 weeks and also with a spatial window of maximum 25 km radius, to assess the impact of these parameters on the number of signals generated.

See Appendix S2 for further details on use and settings of the scan-statistics.

Results

LRI-clusters

Between Feb 1st 1999 and Sept 30th 2006, a total of 35 LRI-clusters with 221 cluster-signals were detected by weekly analysis (Table 1, non-restrictive parameter settings, recurrence interval ≥1 year). By raising the threshold (recurrence interval ≥5 years), we observed only 24 clusters with 146 cluster-signals (respectively 31% and 34% decrease). Figure 2a shows all LRI-clusters and signals on a timescale for the different recurrence interval levels – as detected with the initial non-restrictive parameter settings for space and time windows. The time between the first and the last signal within one cluster ranged from 0 to 26 weeks. By daily analysis, in 1999 and 2006 a total of 194 cluster-signals were detected (compared to 75 signals by weekly analysis with a ≥1 year recurrence level, both with non-restrictive parameter settings). However, the number of clusters was lower (10 clusters by daily analysis vs 12 by weekly analysis in 1999 and 2006).

Figure 2a and Table 1 also show the likely causes for the detected LRI-clusters (according to the criteria in Figure 1, see methods section). The known Legionnaires' disease outbreaks in 1999 and 2006 were detected by LRI-clusters, since cluster-signals were generated with an increased proportion of patient discharge diagnoses for Legionnaires' disease in both outbreak areas and periods (Table 1, Figure 2a and 3a–b) (proportions differed between successive signals: 44–65% in 1999, and 21–63% in 2006; p:<0.0001). The 1999 Legionnaires' disease related cluster-signals included a higher proportion of persons 50–64 years of age (37–48%; p:<0.0001). We compared the earliest detection dates for these outbreaks for daily and weekly analysis. Daily analysis signaled the outbreak 4 days earlier than weekly analysis, 2 days before the national alarm was given during the 1999 Legionnaires' disease outbreak. The 2006 Legionnaires' disease outbreak was detected by weekly analysis on 2006 July 15th, and could have been detected by daily analysis 5 days earlier, 3 days after the national alarm was given.

Many of the other clusters and signals seemed to be related to local RSV and/or influenza activity (70% of cluster-signals and 60% of clusters, Table 1). Some of the influenza and RSV related clusters tended to persist over longer periods (Figure 2a). Young children (0–4 years old) were overrepresented in 82 of the 99 cluster-signals that we scored as RSV related (Table 1; p:<0.05).

In 2000, a cluster was detected with an unusually high number of patients diagnosed with aspergillosis, which was traced to a registration error (one patient was accidentally registered under 28 different anonymous identifiers).

For 46 cluster-signals we did not find a "likely cause" according to the criteria in Figure 1. Of these, 6 belonged to influenza and/or RSV related clusters (Figure 2a), and 11 coincided with local ILI-elevations outside the influenza season (1 at the end of spring 2000, 4 at the end of summer 2000 and 6 at the end of 2005).

When repeating the weekly analyses with restricted time or spatial windows, both Legionnaires' disease outbreaks were still detected with the same timeliness. Table 1 and Figure 2b and 2c also show the clusters and signals that were still detected with a temporal window of maximum 7 weeks, and with a spatial window of maximum 25 km respectively (as compared to the signals detected with the initial non-restrictive settings).

With a time window of maximum 7 weeks, 129 of the 221 initial cluster-signals and 30 of the initial 35 clusters were still detected (respectively 42% and 14% decline, Table 1). Of the 5 clusters not detected — as compared to the initial analyses — 2 had been scored as likely due to RSV, 1 to influenza and for the other 2 no likely cause had been scored (Table 1 and Figure 2a–b).

With a maximum 25 km radius, 165 of the 221 initial cluster-signals and 33 of the 35 clusters were still detected (respectively 25% and 6% decline, Table 1). One of the 2 undetected clusters had been scored as likely due to RSV, for the other no likely cause had been scored (Table 1, Figure 2a and 2c).

Table 1. Detected LRI-clusters and signals between 1999 Feb 1st and 2006 Sept 30th by weekly analysis (recurrence interval ≥1 or ≥5 years) for different parameter settings.

	(A) Non restrictive settings for time and spatial windows				(B) Maximum 7 weeks time window				(C) Maximum radius 25 km			
(Likely) cause	LRI-cluster-signals		LRI-clusters*		LRI-cluster-signals		LRI-clusters*		LRI-cluster-signals		LRI-clusters*	
	Recurrence interval		Recurrence interval		Recurrence interval		Recurrence interval		Recurrence interval		Recurrence interval	
	≥1 yr.	≥5 yr.	≥1 yr.	≥5 yr.	≥1 yr.	≥5 yr.	≥1 yr.	≥5 yr.	≥1 yr.	≥5 yr.	≥1 yr.	≥5 yr.
Legionnaires' disease outbreak 1999	10	10	1	1	7	7	1	1	10	10	1	1
Legionnaires' disease outbreak 2006	4	4	1	1	4	4	1	1	4	4	1	1
Local RSV activity	99	78	9	7	62	55	7	7	68	56	8	6
Local influenza activity	55	28	8	5	25	13	7	5	40	17	9	4
Local RSV *and* influenza activity	n/a	n/a	4	3	n/a	n/a	4	3	n/a	n/a	3	2
Other specific pathogen**	7	6	1	1	5	5	1	1	7	6	1	1
No cause defined***	46	20	11	6	26	12	9	4	36	16	10	5
Total	221	146	35	24	129	96	30	22	165	109	33	20

The total number of detected clusters and signals is presented, for the non-restrictive parameter settings on space and time (A), for the settings with a maximum time window of 7 weeks (B), and for the settings with a maximum radius of 25 km (C). The distribution of (likely) causes according to the criteria in Figure 1 is also presented in the Table.
* A cluster is defined by a set of successive cluster-signals that overlap in space and time.
** The cluster-signals in this category formed only one cluster, which appeared to be caused by a data artifact.
*** Possibly unreported/undetected local LRI-outbreaks by undetected pathogens.
doi:10.1371/journal.pone.0010406.t001

Some of the cluster-signals detected with restrictive time/spatial windows had not been detected with the initially detected signals (data not shown). With the restrictive time window 2 borderline significant cluster-signals were detected, that had been non-significant in the initial analysis. This was due to the fact that the restrictive settings limited the adjustments for taking into account the multiple testing (stemming from the many potential cluster locations and sizes evaluated) [15]. With the restrictive spatial window 3 extra cluster-signals were detected due to the same mechanism, and 2 other extra cluster-signals were detected due to the fact that initial cluster-signals that geographically overlapped with them had dropped out.

Discussion

In this study, prospective surveillance of hospitalization data was simulated using retrospective data, to evaluate whether syndromic surveillance can effectively detect local outbreaks of lower-respiratory infections (LRIs). Over 1999–2006 (400 weeks), 35 space-time LRI-clusters were detected by weekly analysis, with a total of 221 generated cluster-signals. This represents an average rate of approximately 5 new clusters per year, or 3 per year using a threshold recurrence interval ≥5 years. The number of clusters detected per year differed over the study period, reflecting substantial annual variation in influenza epidemics.

Two clusters were related to the Legionnaires' disease "test-case" outbreaks and would have been detected around the same time as the outbreaks were actually detected. This indicates that syndromic surveillance will pick up similar outbreaks of severe respiratory disease in a timely manner. Note that the Legionnaires' disease outbreaks are used here as "positive controls" (or Gold Standard) for realistic severe respiratory outbreaks by uncommon pathogens that may not be (timely) detected by traditional surveillance, such as the Dutch Q-fever outbreak in 2007, for which the initial diagnoses were delayed by several weeks [21,22]. As 17 out of the total 35 LRI clusters probably reflected local RSV and/or influenza activity, many signal investigations could be limited to checking their concurrence with local RSV and/or influenza activity. The 3 clusters with "unknown cause", that concur with local ILI-elevations outside the influenza season, possibly represent very early local influenza activity or local activity of another respiratory pathogen reflected in both GP-ILI-data and hospital LRI-data. For these 3 clusters and the other 8 clusters for which no likely cause was defined, it would have been interesting to investigate possible causes in a truly prospective setting (e.g., by additional diagnostics). Some of these clusters possibly represent unreported and/or undetected local LRI-outbreaks.

As a threshold value for the significance of cluster signals, we used a threshold of recurrence intervals ≥1 year, and only evaluated the LRI clusters that were above this threshold. To illustrate the impact of changing the threshold we repeated the analyses for recurrence intervals ≥5 years. At both threshold levels, two LRI clusters showed a higher proportion of Legionnaires' disease cases (p:<0.0001, see also results section) overlapping with the known outbreak areas, which made us conclude that these LRI clusters indeed detected the Legionnaires' disease outbreaks.

The results of the sensitivity analysis show that the test outbreaks are still detected with the restricted time and spatial windows (at both threshold levels), without loss of timeliness and with less signals generated in time. To limit the computation time we only performed a modest sensitivity analysis. In this study, the restrictions on the time window almost halved the number of signals (42% decline), whereas the clusters in time to investigate declined much less (14% decline). The spatial restrictions resulted in less decline in generated signals (25% decline in signals and 6% decline in clusters). This indicates that with little loss of sensitivity, the restricted time window would be most appropriate to limit the number of generated signals.

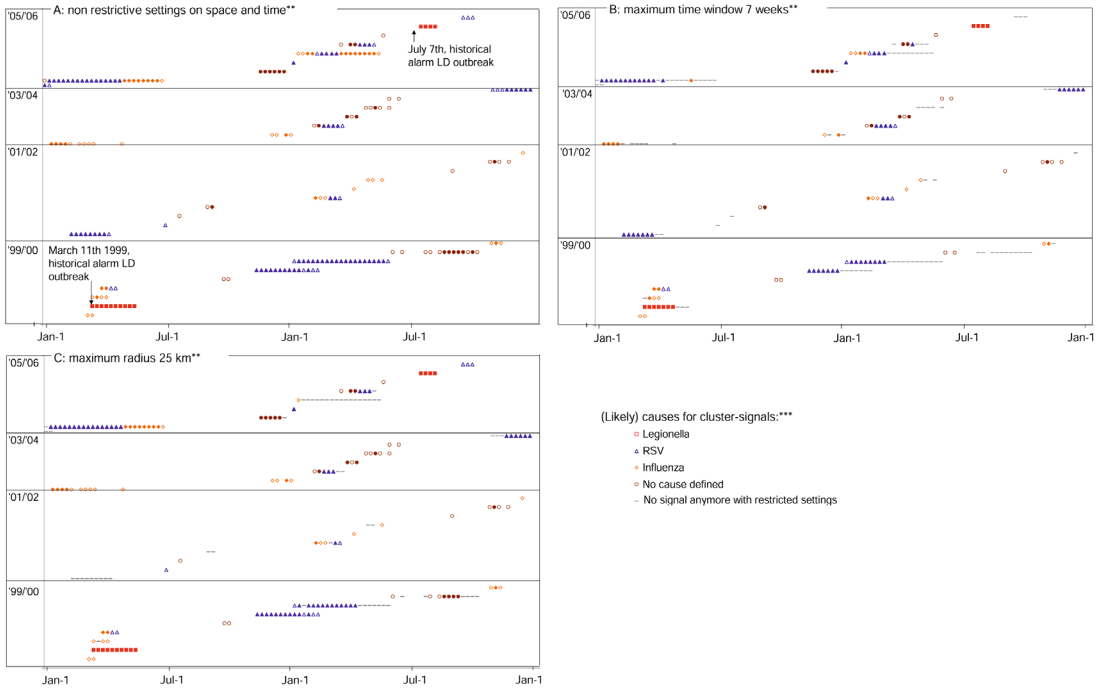


Figure 2. Clusters and generated cluster-signals on a timescale, including all (likely) causes (by weekly analysis).* *Clusters are indicated by sets of successive space-time overlapping cluster-signals placed next to each other on the same height on the y-axis. The cluster-signals caused by a data artifact in 2000 are not presented in the graphs. See Figure 1 for the criteria by which the likely causes were defined and see the Figure 2 legend for the graphic indication of likely causes. **In Figure 2a — for the analyses with non-restrictive settings on time and spatial windows — all detected clusters and signals are presented, as well as the (likely) causes according to the criteria in Figure 1. Figure 2b presents the signals and clusters that are still detected with a maximum time window of 7 weeks, and Figure 2c signals and clusters still detected with a maximum radius of 25 km. ***Signals indicated by open symbols (e.g. "○") have a ≥1 year recurrence interval, coloured symbols (e.g. "●") have a ≥5 yr recurrence interval. A recurrence interval reflects how often a signal of the observed significance level would be observed by chance [18]. I.e. if the recurrence interval of a signal is say 1 year, 1 signal of the observed significance is expected in 1 year.
doi:10.1371/journal.pone.0010406.g002

To our knowledge this is the first study that evaluates the performance of syndromic surveillance with nationwide high coverage data (80–99% of hospitalizations) over a longer period (8 years) with all detected clusters analyzed and (if possible) explained in a systematic way. Feasibility of localized outbreak detection is demonstrated without swamping true signals by excessive false alarms. Some other studies evaluating the performance of space-time syndromic surveillance have concluded differently, but these studies were based on shorter periods, had lower coverage or lacked comparable outbreaks which could be tested [8,23,24]. Cooper et al. tracked the spatial diffusion of influenza and norovirus, using space-time analysis on syndromic data from a telephone help line system in the UK, but did not test space-time detection for more localized outbreaks [23]. Using syndromic surveillance for detection of local gastro-intestinal outbreaks in New York City, Balter et al. found numerous cluster-signals in time, but these could not be used for effective surveillance because of insufficient comparable diagnostic data [8]. Respiratory disease outbreaks could not be evaluated in the NYC study, because no local respiratory outbreaks had been reported in the study period. Nordin et al. used simulated anthrax attack data injected in true physician's visit data to confirm that a respiratory outbreak initiated by bioterrorism will be detected in a timely manner by syndromic surveillance [24]. However, no results on the number of possibly false alarms were presented. These studies present space-time cluster detection analyses over relatively few years and are therefore prone to miss the effects of annual variation. Furthermore, sensitivity for local outbreaks is reduced by using data with relatively low coverage levels. For such data sources with low coverage, methods other than space-time scan statistics seem more appropriate to generate useful information for public health practice (like aberration detection in time).

We performed weekly analyses (instead of daily) over the whole study period, because these analyses consume considerable computation time. Daily analyses in 1999 and 2006 detected fewer clusters than weekly analyses because the threshold level for recurrence intervals (≥1 year) is more strict (see Appendix S2). Daily analyses would therefore probably not detect more epidemiological events but would yield more timely signals.

Hospital based syndromic surveillance could be a helpful tool in detecting local LRI-outbreaks, complementing outbreak detection by laboratory surveillance or astute clinicians. Syndromic surveillance might be most valuable for outbreaks due to uncommon or novel pathogens (like the SARS outbreak), as these seem more likely to be missed by the laboratory and clinicians. Furthermore, outbreaks due to more common pathogens could also be missed, as for community acquired pneumonia often no causative pathogen is detected [25,26]. Apart from that, under-notification can complicate outbreak detection through laboratories and clinicians [20].

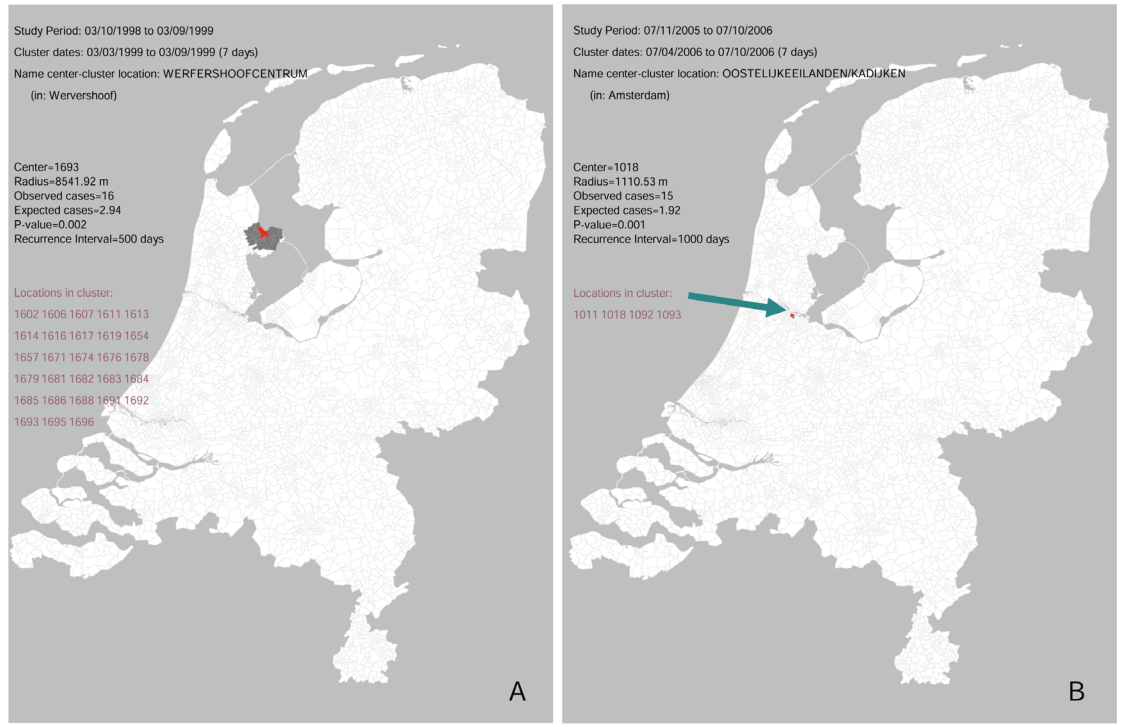


Figure 3. The earliest detected Legionnaires' disease outbreak related LRI-cluster-signals (1999 and 2006) as presented on a map of the Netherlands (by daily analysis). Figure 3a and 3b show the cluster-signals that detected the 1999 and 2006 outbreak respectively. Output of the Satscan scan-statistic software is presented in the legend. On the map the borders of all postal code areas are indicated, the postal code areas of the cluster-signals are marked in dark-grey with the center postal code marked in red.
doi:10.1371/journal.pone.0010406.g003

A prerequisite for prospective syndrome surveillance is the real-time availability of hospitalization data, including clinical diagnoses and symptoms by date of hospitalization. Although at present not available in the Netherlands, such real-time syndromic data collection may become feasible after the nationwide implementation of electronic health-care information exchange. In this light, the results of our study justify further development of these methods, including retrospective evaluation of other types of documented health events than the ones presented in our study.

Besides that, further research should focus on prospective application of these methods. In a prospective setting, sustaining reliable data with high coverage and few data artifacts might be more challenging, thus possibly leading to higher numbers of false alarms. In addition, it should be evaluated to what extent 3 to 5 new syndromic clusters per year would indeed be manageable in a prospective setting. Responding to such clusters is complicated, because the cause and thus possible threat will initially often be unknown. For each new cluster, it should first be verified whether plausible explanations can be found in epidemiological or laboratory data. For example, LRI clusters need to be interpreted in relation to local influenza or RSV activity similar as we did in our study, and provided the age distribution of cases reflects the usual pattern, further investigation would seem unnecessary. Internet-based ILI-monitoring [27] combined with virological self -sampling (at home) [28] could increase the microbiological base for interpreting syndromic surveillance data. Age stratified syndromic surveillance with a multivariate space-time scan statistic [29] may further facilitate quick interpretation of clusters by revealing the affected age groups.

Conclusion

This retrospective study shows that space-time syndromic surveillance on hospitalizations can timely detect local LRI-outbreaks independent of detection of the causative pathogen. The frequency of cluster detection, when interpreted in the light of available epidemiological and microbiological data, does not give rise to excessive levels of further investigations.

Consequently, we recommend real-time syndromic surveillance as an additional tool for detection of local LRI outbreaks, but only if syndromic data with sufficient quality and coverage can be collected, coupled with epidemiological and microbiological data. Public health responses can be based on a combination of syndromic surveillance data, reports by astute clinicians and early diagnostic test results, which all could generate the first alarm for different kinds of disease events. Future research on prospective syndromic surveillance should therefore focus on practical methods for integrating syndromic surveillance alarms with clinical reports and laboratory information for effective public-health responses.

Supporting Information

Appendix S1 Detailed syndrome definition for hospitalizations with lower-respiratory infection syndrome.

Found at: doi:10.1371/journal.pone.0010406.s001 (0.06 MB DOC)

Appendix S2 Details on space-time analyses and Satscan settings.
Found at: doi:10.1371/journal.pone.0010406.s002 (0.03 MB DOC)

Acknowledgments

We thank Dutch Hospital Data for providing data from the Dutch National Medical Register (LMR) and we thank Douglas Fleming for reading and commenting on the manuscript.

References

1. Buehler JW, Berkelman RL, Hartley DM, Peters CJ (2003) Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 9: 1197–1204.
2. Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R (2001) Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. BMC Public Health 1: 9.
3. Fleming DM, Barley MA, Chapman RS (2004) Surveillance of the bioterrorist threat: a primary care response. Commun Dis Public Health 7: 68–72.
4. Miller M, Roche P, Spencer J, Deeble M (2004) Evaluation of Australia's National Notifiable Disease Surveillance System. Commun Dis Intell 28: 311–323.
5. Ohkusa Y, Shigematsu M, Taniguchi K, Okabe N (2005) Experimental surveillance using data on sales of over-the-counter medications–Japan, November 2003–April 2004. MMWR Morb Mortal Wkly Rep 54 Suppl: 47–52.
6. Heffernan R, Mostashari F, Das D, Karpati A, Kuldorff M, et al. (2004) Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 10: 858–864.
7. Lai PC, Wong CM, Hedley AJ, Lo SV, Leung PY, et al. (2004) Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. Environ Health Perspect 112: 1550–1556.
8. Balter S, Weiss D, Hanson H, Reddy V, Das D, et al. (2005) Three years of emergency department gastrointestinal syndromic surveillance in New York City: what have we found? MMWR Morb Mortal Wkly Rep 54 Suppl: 175–180.
9. Hope K, Durrheim DN, d'Espaignet ET, Dalton C (2006) Syndromic Surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 60: 374–375.
10. Olson KL, Grannis SJ, Mandl KD (2006) Privacy Protection Versus Cluster Detection in Spatial Epidemiology. Am J Public Health 96: 2002–2008.
11. Higgs BW, Mohtashemi M, Grinsdale J, Kawamura LM (2007) Early detection of tuberculosis outbreaks among the San Francisco homeless: trade-offs between spatial resolution and temporal scale. PLoS ONE 2: e1284.
12. Donker GA [Continuous morbidity registration sentinels Netherlands 2006]. NIVEL, annual report, 2007. Available: http://www.nivel.nl/pdf/CMR-Peilstations-2006.pdf. Accessed 2010 Apr 8.
13. Den Boer JW, Yzerman EP, Schellekens J, Lettinga KD, Boshuizen HC, et al. (2002) A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. Emerg Infect Dis 8: 37–43.
14. Van den Hoek JA, IJzerman EP, Coutinho RA (2006) [Legionella outbreak in Amsterdam: a cooling tower as the source]. Ned Tijdschr Geneeskd 150: 1808–1811.
15. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F (2005) A Space-Time Permutation Scan Statistic for Disease Outbreak Detection. PLoS Med 2: e59.
16. Kulldorff M, Information Management Services, Inc. (2006) SaTScanTM v7.0: Software for the spatial and space-time scan statistics. Available: http://www.satscan.org. Accessed 2008 Mar 18.
17. Abrams AM, Kleinman KP (2007) A SaTScan macro accessory for cartography (SMAC) package implemented with SAS software. Int J Health Geogr 6: 6.
18. Kleinman K, Lazarus R, Platt R (2004) A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. Am J Epidemiol 159: 217–224.
19. Slobbe LCJ, de Bruin A, Westert GP, Kardaun JWPF, Verweij GCG, [Classification of diagnoses and procedures and application in new hospital statistics]. Bilthoven, the Netherlands, 260201002: RIVM; 260201002: Available: http://www.rivm.nl/bibliotheek/rapporten/260201002.pdf. Accessed 2010 Apr 8.
20. Van Hest NA, Hoebe CJ, den Boer JW, Vermunt JK, IJzerman EP, et al. (2007) Incidence and completeness of notification of Legionnaires' disease in The Netherlands: covariate capture-recapture analysis acknowledging regional differences. Epidemiol Infect 136: 540–550.
21. Van Steenbergen JE, Morroy G, Groot CA, Ruikes FG, Marcelis JH, et al. (2007) [An outbreak of Q fever in The Netherlands–possible link to goats]. Ned Tijdschr Geneeskd 151: 1998–2003.
22. Morroy G, Schimmer B, Duynhoven Y, Wijkmans C (2008) [Q-fever: will the mystery be solved?]. Infectieziekten Bulletin 19: 66–68. Available: http://www.rivm.nl/cib/binaries/IB_1902-p66-68_tcm92-49568.pdf. Accessed 2010 Apr 8.
23. Cooper DL, Smith GE, Regan M, Large S, Groenewegen P (2008) Tracking the spatial diffusion of influenza and norovirus using telehealth data: a spatio-temporal analysis of syndromic data. BMC Med 6: 16.
24. Nordin JD, Goodman MJ, Kulldorff M, Ritzwoller DP, Abrams AM, et al. (2005) Simulated anthrax attacks and syndromic surveillance. Emerg Infect Dis 11: 1394–1398.
25. Templeton KE, Scheltinga SA, van den Eeden WC, Graffelman AW, van den Broek PJ, et al. (2005) Improved diagnosis of the etiology of community-acquired pneumonia with real-time polymerase chain reaction. Clin Infect Dis 41: 345–351.
26. Van der Eerden MM, Vlaspolder F, de Graaff CS, Groot T, Jansen HM, et al. (2005) Value of intensive diagnostic microbiological investigation in low- and high-risk patients with community-acquired pneumonia. Eur J Clin Microbiol Infect Dis 24: 241–249.
27. Marquet RL, Bartelds AI, van Noort SP, Koppeschaar CE, Paget J, et al. (2006) Internet-based monitoring of influenza-like illness (ILI) in the general population of the Netherlands during the 2003–2004 influenza season. BMC Public Health 6: 242.
28. Cooper DL, Smith GE, Chinemana F, Joseph C, Loveridge P, et al. (2008) Linking syndromic surveillance with virological self-sampling. Epidemiol Infect 136: 222–224.
29. Kulldorff M, Mostashari F, Duczmal L, Katherine Yih W, Kleinman K, et al. (2007) Multivariate scan statistics for disease surveillance. Stat Med 26: 1824–1833.
30. Kulldorff M (2001) Prospective time periodic geographical disease surveillance using a scan statistic. J R Statist Soc A 164: 61–72.

Author Contributions

Conceived and designed the experiments: CCvdW LvA WvP NJDN MPGK. Performed the experiments: CCvdW. Analyzed the data: CCvdW. Contributed reagents/materials/analysis tools: GD GAD. Wrote the paper: CCvdW. Helped drafting and reviewed drafts of the paper: LvA WvP NJDN GAD WvdH MPGK. Helped interpreting the analysis results: LvA WvP NJDN MPGK. Reviewed drafts of the paper: GD.

Chapter 5

In Search of Hidden Q-fever Outbreaks: Linking Syndromic Hospital Clusters to Infected Goat Farms

Epidemiolog Infect 2010 [Epub ahead of print]

In search of hidden Q-fever outbreaks: linking syndromic hospital clusters to infected goat farms

C. C. VAN DEN WIJNGAARD[1]*, F. DIJKSTRA[1], W. VAN PELT[1], L. VAN ASTEN[1], M. KRETZSCHMAR[1,2], B. SCHIMMER[1], N. J. D. NAGELKERKE[3], P. VELLEMA[4], G. A. DONKER[5] AND M. P. G. KOOPMANS[1,6]

[1] *National Institute for Public Health and the Environment, Bilthoven, The Netherlands*
[2] *Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, The Netherlands*
[3] *United Arab Emirates University, Al-Ain, United Arab Emirates*
[4] *Animal Health Service, Deventer, The Netherlands*
[5] *Nivel, Netherlands Institute of Health Services Research, Utrecht, The Netherlands*
[6] *Erasmus Medical Center, Rotterdam, The Netherlands*

(*Accepted 14 April 2010*)

SUMMARY

Large Q-fever outbreaks were reported in The Netherlands from May 2007 to 2009, with dairy-goat farms as the putative source. Since Q-fever outbreaks at such farms were first reported in 2005, we explored whether there was evidence of human outbreaks before May 2007. Space–time scan statistics were used to look for clusters of lower-respiratory infections (LRIs), hepatitis, and/or endocarditis in hospitalizations, 2005–2007. We assessed whether these were plausibly caused by Q fever, using patients' age, discharge diagnoses, indications for other causes, and overlap with reported Q fever in goats/humans. For seven detected LRI clusters and one hepatitis cluster, we considered Q fever a plausible cause. One of these clusters reflected the recognized May 2007 outbreak. Real-time syndromic surveillance would have detected four of the other clusters in 2007, one in 2006 and two in 2005, which might have resulted in detection of Q-fever outbreaks up to 2 years earlier.

Key words: Outbreaks, Q fever, respiratory infections, surveillance, zoonoses.

INTRODUCTION

Q fever is a zoonosis occurring worldwide caused by *Coxiella burnetii*, an intracellular bacterium. Although most human infections remain asymptomatic or present as a non-specific flu-like illness, severe acute Q fever presents primarily with atypical pneumonia or hepatitis. The infection poses an increased risk for pregnant women and persons with heart-valve disorders or impaired immunity, who may develop chronic disease with endocarditis as its most frequent chronic clinical manifestation. Q fever in pregnancy, whether symptomatic or asymptomatic, may result in adverse pregnancy outcomes. Cattle, sheep and particularly goats are considered the primary reservoirs from which human infection occurs, typically by inhalation of infected aerosols, and less commonly through ingestion [1–3].

For many years, Q fever in humans was very rare in The Netherlands, with around 15 reported cases per year [4], but since the end of May 2007, outbreaks have occurred in rural areas mainly in the southern part of the country, with 20–25% of the reported cases requiring hospitalization [5]. In 2007, 178 Q-fever cases were reported, most of which occurred in May and June 2007. In total, 31% ($n=55$) of these cases occurred in a relatively small rural area [6]. In 2008 and 2009, large outbreaks of Q fever recurred with increasing numbers of reported cases and an expanding geographic area [5, 7].

Dairy-goat farms are considered the most likely source of infection for these outbreaks, although evidence is still inconclusive [5, 7, 8]. Q-fever abortion waves have been reported at several dairy-goat farms, starting at least 2 years before the first recognized human outbreak [5, 9]. This time lag raised the question whether unrecognized human outbreaks may have preceded May 2007, particularly since most severe cases present as pneumonia, for which laboratory tests are often not requested. In fact, the diagnoses of the May/June 2007 outbreak cases were delayed by several weeks, until pneumonia patients were retested for *C. burnetii*, triggered by an increase in the number of Q fever reports in humans in the region [8, 10].

Syndromic surveillance allows monitoring of clinical syndromes such as 'pneumonia' or 'lower-respiratory infection' (LRI) independent of laboratory confirmation. It also permits detection of outbreaks of diseases for which diagnostics are either not available or not requested. In order to optimize its sensitivity for detection of local outbreaks, surveillance of syndrome spikes can be performed in space and time [11, 12]. Earlier we demonstrated the value of this approach by showing that syndromic surveillance of Dutch hospitalizations can indeed detect local lower-respiratory disease outbreaks [13].

In the current study, we retrospectively explored whether there is evidence for human Q-fever outbreaks in The Netherlands in 2005–2007, before the May/June 2007 outbreak. Using space–time scan statistics, we looked for local increased numbers of hospitalized patients with LRIs and other syndromes that can be associated with *C. burnetii* infection. Based on available epidemiological and surveillance data, and the geographical proximity of small-ruminant farms that tested positive for Q fever, we then assessed whether these local increases could have been caused by Q fever or by other infections like respiratory syncytial virus (RSV) or influenza. Finally, we evaluated whether real-time syndromic surveillance of hospitalizations could have accelerated the detection of human Q-fever outbreaks.

METHODS

Hospitalization data

For the period 1 January 2005 to 30 September 2007, hospitalization data were obtained from the Dutch National Medical Register, which has about 80% population coverage. We excluded data before 2005, as data on farms that tested positive for Q fever was not then available. The data included discharge and secondary diagnoses by date of hospitalization. We selected all hospitalization records showing diagnoses involving clinical syndromes compatible with Q fever, i.e. LRI, hepatitis, and/or endocarditis. We used a case definition for LRI, which has proved functional for detecting severe respiratory disease outbreaks in hospitalization data [13]. See Appendix A (available online) for detailed hepatitis and endocarditis case definitions.

We analysed 108 338 hospitalizations for LRI after excluding 5% of 114 245 records because of patients readmitted within six consecutive weeks; 3826 hospitalizations for hepatitis after excluding 29% of 5382 records because of readmissions from 1999 to 2007 (to avoid chronic cases first hospitalized before 2005); and 2130 hospitalizations for endocarditis after excluding 18% of 2612 records because of readmissions from 1999 to 2007 (to avoid chronic cases first hospitalized before 2005). Data were aggregated by week of hospitalization, postal code, and patients' age group (0–4, 5–19, 20–49, 50–64, $\geqslant$65 years). Of these, we regarded the 20–49 and 50–64 years age groups to be at higher risk for Q fever, since most reported cases were adults, with the median age $\sim$50 years [5, 7]. Since higher levels of spatial resolution can result in more sensitive detection of outbreaks [14, 15], we used 4-digit postal codes designating 4023 areas in a 16·3 million population. This provided the highest level of spatial resolution available within national privacy regulations.

Infectious disease surveillance data

To assess alternative causes (other than Q fever) for detected local clusters of syndromic hospital cases, we explored mandatory reports of psittacosis, Legionnaires' disease, hepatitis A, B and C, and Q fever, although prior to 2007 Q-fever cases may have remained undetected or misdiagnosed. The counts of

* Author for correspondence: C. C. van den Wijngaard, M.Sc., RIVM, Centre for Infectious Disease Control Netherlands, PO Box 1, 3720 BA Bilthoven, The Netherlands.
(Email: kees.van.den.wijngaard@rivm.nl)

reports were aggregated by week of disease onset and 4-digit postal codes. We used regional data on influenza-like illness (ILI) to assess whether a rise in local LRI cases could be due to influenza. These data are collected weekly by a sentinel network of general practitioners, the Continuous Morbidity Registration (CMR) centres, which provide 1% population coverage, representative by age, gender, geographic distribution and population density [16]. ILI is defined as an acute beginning of a respiratory infection with fever (rectal temperature $\geqslant 38\ ^{\circ}C$), and with at least one of the following symptoms present: cough, coryza, sore throat, frontal headache, retrosternal pain, myalgia. Due to the limited number of sentinel practices, these weekly data were aggregated into four major regions instead of 4-digit postal codes, as for the hospitalization data.

Data on positive goat and sheep farms

To identify possible point sources of *C. burnetii*, we used Q-fever-related abortion waves on dairy-goat and dairy-sheep farms as reported to the Dutch Animal Health Service. For each farm that tested positive for *C. burnetii* between 2005 and 2008, we recorded its 4-digit postal code and the date of its first submission of placental tissue.

Diagnostic tests for Q fever as a cause of abortion waves in animals were first available in 2004. In 2005 the first positive results were obtained from two dairy-goat farms, in 2006 from six dairy-goat farms and one dairy-sheep farm, in 2007 from seven dairy-goat farms, and in 2008 from eight dairy-goat farms and one dairy-sheep farm. Until 12 June 2008, reporting of abortion waves and subsequent laboratory testing of placental tissue were voluntary in The Netherlands; thus previous Q-fever abortion waves cannot be excluded in areas making no reports. For one farm, placental tissue from 2001 was retrospectively tested and found positive for *C. burnetii*.

We regarded these farms as possible point sources of *C. burnetii* during the entire study period, as some were known to have had abortion problems in the year before and/or after they first submitted placental tissue. Moreover, infected animals can become long-term shedders, and *C. burnetii* is very persistent in the environment [17].

Scan statistics for space–time clusters

In order to detect a possible increase in Q-fever-related regional cases in the hospitalization data, we used a space–time permutation scan statistic that compared observed and expected numbers of cases in flexible circular areas over flexible time periods [18]. SaTScan software and SAS (version 9.1, SAS Institute Inc., USA) were used to run the scan-statistic [19, 20].

For each week in the analysis period, we ran the scan statistic on data from the preceding year, thus simulating prospective (real-time) surveillance. This generated weekly space–time signals. To indicate the significance of detected space–time signals, we used recurrence intervals, which indicate how often a signal of the observed significance would be observed by chance under the hypothesis of no outbreak [21]. That is, if the recurrence interval of a signal is say 1 year, one signal of the observed significance is expected in 1 year. If the recurrence interval of a signal was $\geqslant 1$ year, it was viewed as a significant signal. Besides that, we also used a threshold of recurrence intervals $\geqslant 5$ years to indicate highly significant signals. We defined a cluster as a group of successive overlapping space–time signals, since overlap suggests the same cause; we defined a cluster signal as an individual space–time signal within a cluster. Unlike an earlier study [13], in the present study we also assessed whether the same cause for successive overlapping cluster signals seemed unlikely due to shifts in space or time windows or shifts in predominant age groups or discharge diagnoses. We used previously validated parameter settings for the scan statistic, with minimal constraints on the maximum space and/or time windows, to avoid incorrect assumptions about time, geographical location, or size of possible Q-fever outbreaks [13].

We also performed space–time scans on specific infectious disease data (mandatory reports and ILI sentinel data), to assess whether these might explain the space–time clusters of hospitalizations that we detected. For the mandatory reports, the same parameter settings were used as for the hospitalizations. For the ILI data, the space–time scans were performed using both case data and population-at-risk data [22], as previously described [13].

LRI clusters

To determine the LRI clusters for which Q fever was an unlikely, possible or plausible cause, we used two-step criteria (Fig. 1). In step 1, the upper box shows the criteria for LRI clusters unlikely to be caused by Q fever, based on indications for other causative pathogens and epidemiological characteristics (e.g. patients' age) that differed from the confirmed Q-fever cases. The lower box shows the criteria for LRI clusters possibly caused by Q fever, based on epidemiological data [5, 7, 23] as well as overlap with clusters of human Q-fever reports. In step 2, we assessed whether clusters identified as possibly caused by Q fever (in step 1) showed geographical overlap with farms that tested positive in 2005–2008. If so, we considered Q fever a plausible cause for these clusters.

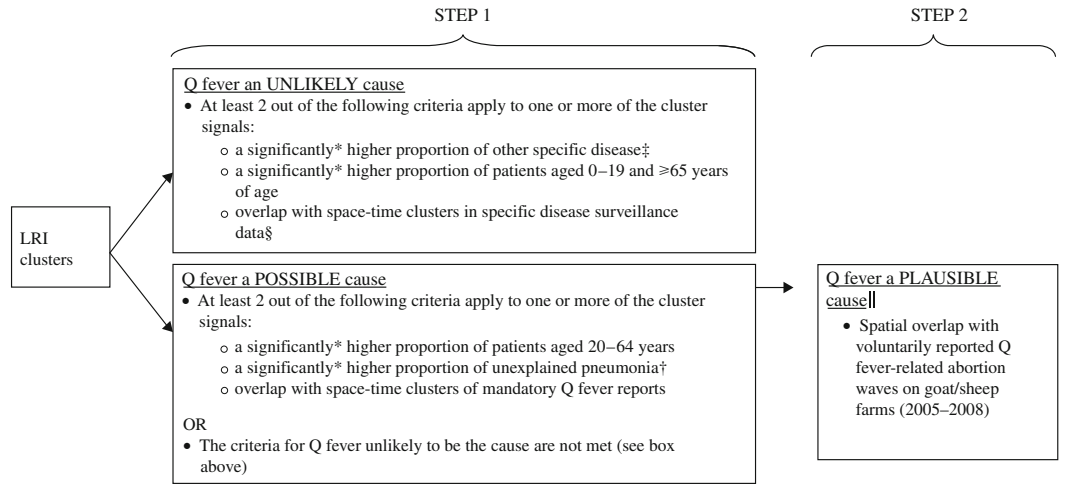


Fig. 1. Two-step criteria to explore the plausibility that Q fever caused the lower-respiratory infection (LRI) hospitalization clusters detected in 2005–2007. * Right-sided Fisher's exact test for 2×2 tables ($\alpha \leqslant 0{\cdot}05$ and/or $< 0{\cdot}01$) of hospitalizations inside *vs.* outside of the cluster signal. The proportion of hospitalizations with a specific disease characteristic (e.g. Legionnaires' disease as discharge diagnoses, or patients aged 20–64 years) can be significantly higher in hospitalizations within the cluster signal than the proportion outside the cluster signal. † ICD-9 codes 485/486/481/4829. ‡ We considered high proportions of bronchitis/bronchiolitis (ICD-9 codes 4801/4660/4661/490) as a likely indication for RSV activity. § Mandatory reports and influenza-like illness sentinel data. || Only assessed if the cluster meets the criteria for Q fever to be a possible cause.

Hepatitis and endocarditis clusters

To assess whether detected hepatitis and/or endocarditis clusters might have been caused by Q fever, we again used two-step criteria, similar as for the LRI clusters. In step 1, we now evaluated whether hepatitis clusters overlapped with clusters of hepatitis A, B, or C reports and assessed whether hepatitis and/or endocarditis clusters had any characteristics (age distribution, specific diagnoses) suggesting causes other than Q fever. In step 2, clusters for which no alternative causes were found, were assessed for overlap with reported Q-fever abortion waves at farms in 2005–2008. If they overlapped, we considered Q fever a plausible cause for these human disease clusters.

The patients of all clusters for which we considered Q fever a plausible cause were then line-listed and inspected for data anomalies that might have caused a cluster signal. Finally, we evaluated to what extent actual prospective syndromic surveillance could have accelerated the detection of human Q-fever outbreaks.

RESULTS

LRI clusters

We detected 20 LRI clusters in hospitalization data between 2005 and 2007 [for details see Appendix B, Table B1 (available online)]. Applying the criteria in Figure 1, for seven of these LRI clusters, Q fever was considered a plausible cause; for three clusters a possible cause, and for ten clusters an unlikely cause. Figure 2 shows all LRI clusters and signals on a timescale, and Figure 3*a* shows the locations of the LRI clusters for which we considered Q fever a plausible cause.

Q fever: a plausible cause

Of the seven clusters for which Q fever was considered a plausible cause, one reflected the known May/June 2007 outbreak, as it strongly overlapped in space and time with that outbreak. If monitored in real time, this

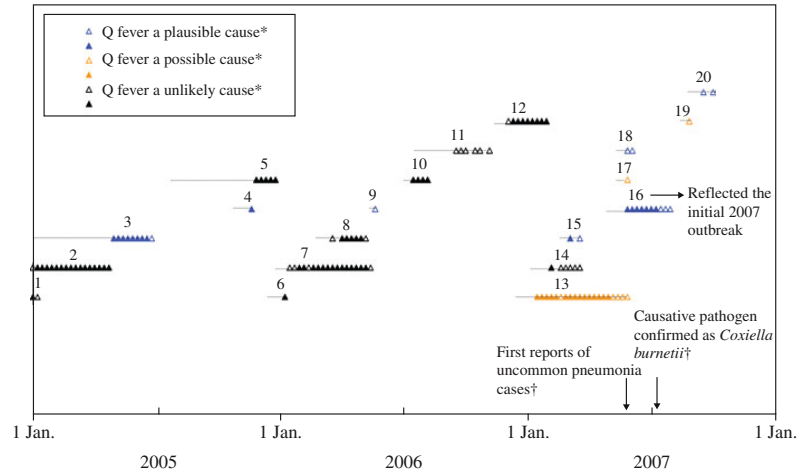


Fig. 2. All lower-respiratory infection (LRI) signals and clusters on a timescale. The legend indicates which clusters have Q fever as a plausible, possible, or unlikely cause (see criteria in Fig. 1). * For each cluster, horizontal dotted lines indicate the total cluster episode, and triangles indicate weekly generated cluster signals. As some clusters overlapped in time, consecutive clusters are presented at different heights at the *y*-axis. The significance level of cluster signals, as measured by the recurrence interval (see Methods section), is indicated by the colour value of the triangles. For those with uncoloured backgrounds, the recurrence interval of the signal is $\geqslant$1 year; with coloured backgrounds, the recurrence interval is $\geqslant$5 years). † At the end of May 2007 the first uncommon pneumonia patients were reported in the area of the initial 2007 outbreak, and on 11 July *C. burnetii* was confirmed to be the causative pathogen. This initial 2007 outbreak was reflected by cluster 16.

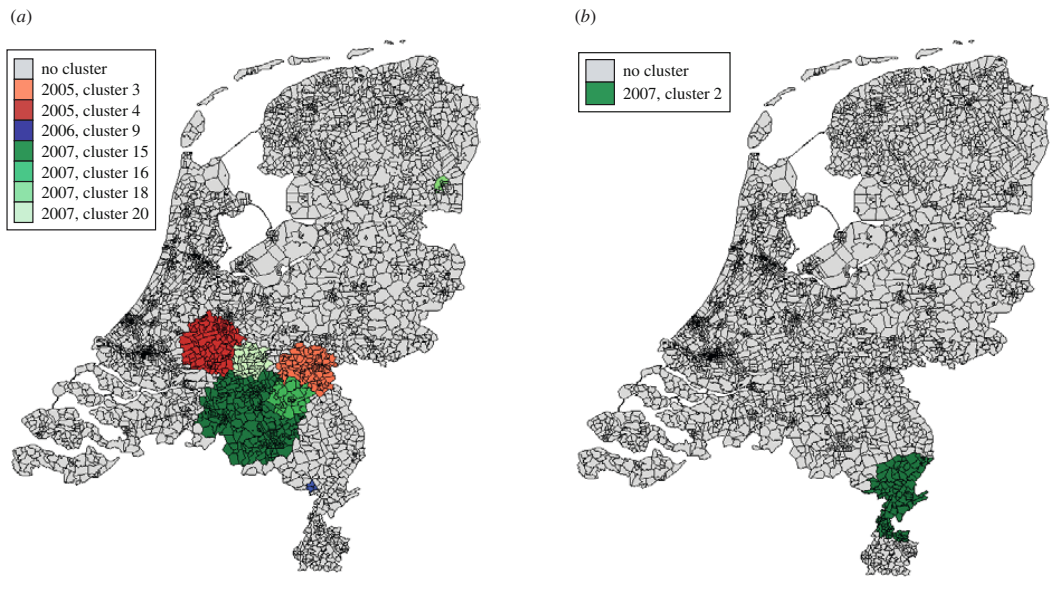


Fig. 3. (*a*) Lower-respiratory infection (LRI) clusters for which Q fever seemed a plausible cause, presented on the 4-digit postal code map of The Netherlands. See criteria for Q fever as a plausible cause in Figure 1. If LRI cluster areas overlapped, the smaller cluster area was drawn on top of the larger area. (*b*) The hepatitis cluster for which Q fever seemed a plausible cause, presented on the 4-digit postal code map of The Netherlands.

cluster would have given the first signal at the end of May 2007, i.e. in the same week that suspicious pneumonia patients were first reported, and 6 weeks before the first positive laboratory results for *C. burnetii* were obtained [8, 10].

Of the other six clusters, three would have given the first signal 1–2 years earlier than the known outbreak in 2007 (two in 2005 and one in 2006), one cluster 3 months earlier, one at the same time, and one cluster 4 months afterwards [Fig. 2 and Appendix B, Table B1 (online)]. All but one of these clusters occurred in the southeast or the middle of the country (Fig. 3*a*) but in a more widespread area than the May/June 2007 outbreak. The two 2005 clusters and two clusters in 2007 were highly significant with recurrence intervals $\geqslant$5 years.

Q fever: a possible cause

For only three clusters was Q fever a possible cause, i.e. not plausible because they lacked geographical proximity to Q-fever abortion waves on farms. All three clusters occurred in 2007 [see Fig. 2 and Appendix B (online)].

Q fever: an unlikely cause

The ten clusters for which Q fever was considered an unlikely cause were not assessed for geographical overlap with infected farms, because other pathogens (e.g. influenza and RSV) seemed probable. As previously described, one cluster was clearly due to *Legionella*, as it comprised many cases of Legionnaires' disease and overlapped in space and time with a known *Legionella* outbreak [13, 24].

Hepatitis and endocarditis clusters

Two hepatitis clusters (two signals) were detected in 2005–2007. The first was possibly due to hepatitis C [Appendix B, Table B2 (online)]. The other overlapped with reported abortion waves on goat farms, making Q fever its plausible cause. The cluster occurred more to the southeast (Fig. 3*b*) than the May/June 2007 outbreak. It would have given the first signal 2 weeks before the first reports of suspicious pneumonia patients in 2007. The cluster does not exceed the threshold level for highly significant clusters (i.e. recurrence interval $\geqslant$5 years).

For endocarditis, we revealed probable duplicates in all three detected clusters based on line-listing the age, postal code, and diagnoses of cluster patients. Although duplicates had been excluded from the analyses by using anonymous patient identifiers, some patients appeared to have obtained two or more anonymous identifiers through transfers between hospitals. We therefore excluded the endocarditis clusters from further analysis.

DISCUSSION

Using space–time syndromic surveillance methods, we found substantial support for the occurrence of human Q-fever outbreaks in The Netherlands preceding the 2007 outbreak and covering a wider area. In proximity to infected small-ruminant farms, local clusters of human LRI and hepatitis were detected in hospitalization data for 2005–2007. Although the retrospective nature of this study precluded laboratory confirmation, available epidemiological and surveillance data suggested Q fever as the cause for several syndrome clusters that mostly occurred in the southeast and middle of the country.

Prospective syndromic surveillance on hospitalizations, as simulated in this study, would have signalled some of these clusters long before detection of the May/June 2007 outbreak. The space–time signals of LRI or hepatitis could then have prompted further evaluation and/or laboratory tests, possibly confirming *C. burnetii* to be the causative pathogen. Although such efforts might not have prevented the large-scale outbreaks in the following years, it possibly would have lead to the detection and the appropriate treatment of Q-fever patients in a more timely manner. Apart from that, earlier detection of human Q-fever outbreaks might have facilitated research into transmission routes between specific farms and humans. For the Dutch human Q-fever outbreaks after 2007, this research has been complicated by dissemination of disease in the dairy-goat population, leading to widespread environmental contamination. These factors complicate the identification of specific farms as the source of human infections.

Of the LRI clusters detected in this study, one reflected the already known 2007 outbreak, confirming that syndromic surveillance can indeed detect Q-fever clusters, consistent with our findings on syndromic detection of respiratory outbreaks [13]. For six other LRI clusters and one hepatitis cluster as well, Q fever seemed a plausible cause. So far, hepatitis due to Q fever has been reported sporadically in The Netherlands: 33 hepatitis cases out of 1000 reported Q-fever cases in 2008 [5] and 5/178 in 2007. Hepatitis due to Q

fever might be more likely from infection by ingestion rather than inhalation [25, 26] and may thus follow distribution channels of contaminated unpasteurized products like raw cheese. If such contaminated products were to cause infections nationwide (without regional clustering), this would not be detected by our space–time analysis.

Our results suggest that the rise in reported Q-fever cases in 2008 and 2009 may partially reflect increased awareness among clinicians following the first recognized outbreak in 2007. This would be consistent with the fact that in 2008 and 2009, Q-fever reports originated from a wider area that overlaps with the area of most hospitalization clusters of 2005–2007 for which Q fever seemed a plausible cause. Moreover, the proportion of hospitalizations in confirmed Q-fever cases was smaller in 2008 and 2009 than in 2007, possibly indicating that increased awareness among clinicians led to diagnosis of earlier and milder Q-fever infections [5]. However, in 2007 we observed four clusters for which Q fever seemed a plausible cause *vs*. a total of three clusters in 2005 and 2006 combined, which suggests an actual rise in Q-fever infections in 2007.

Our finding that syndromic hospital data can reveal disease clusters possibly caused by Q fever, supports the value of prospective syndromic surveillance for detection of otherwise hidden outbreaks. We showed previously that syndromic surveillance for LRI would detect a modest number of clusters in time to investigate: on average five clusters per year at the lowest threshold level (recurrence intervals $\geqslant 1$ year), and three clusters at the highest (recurrence intervals $\geqslant 5$ years) [13]. If such prospective surveillance were to focus on detection of Q-fever outbreaks, inclusion of data on infected small-ruminant farms, as in the current study, would further decrease the number of clusters to investigate. Nevertheless, even then such surveillance also requires sufficient complementary epidemiological and microbiological data to guide further investigation of detected clusters. For example, LRI clusters are best interpreted in relation to local influenza or RSV surveillance data. A prerequisite for prospective syndrome surveillance is the real-time availability of hospitalization data, including clinical diagnoses and symptoms by date of hospitalization. Although at present not available in The Netherlands, such real-time syndromic data collection might become feasible due to the nationwide implementation of electronic healthcare-information exchange. Finally, proper data collection and analysis will be more challenging in a prospective setting, for example due to data quality problems such as reporting delays.

This study has some further limitations. Due to its retrospective setting, *C. burnetii* infection could not be laboratory-confirmed for detected cluster patients. Therefore, we could only indicate whether Q fever seemed a plausible cause by excluding clusters with other apparent causes and assessing the presence of *C. burnetii* in cluster areas, as measured by overlap with infected farms. Another limitation lies in our use of voluntary reports from farmers to indicate possible *C. burnetii* point sources. Cluster areas not overlapping with reportedly infected farms may still have been contaminated by farms not tested or, possibly, by the spreading of manure from non-local contaminated farms. If so, this may explain those clusters that seemed possibly due to Q fever but did not overlap with reported Q fever on farms.

Serological testing of preserved human samples might confirm the occurrence of Q-fever clusters before 2007. However, to our knowledge no historical samples from specific LRI patients within the cluster areas are available. Other studies have found that substantial proportions of a human population can be exposed to Q fever without symptomatic infections [27, 28]. Therefore, a solution might be to use historical samples from blood donors, if sufficient samples from the cluster areas are still available.

This study shows substantial support for the occurrence of human Q-fever outbreaks in The Netherlands before detection of the May/June 2007 outbreak and covering a wider area. Retrospectively, suspicious LRI and hepatitis hospitalization clusters from 2005 to 2007 were detected and found to overlap with small-ruminant farms reporting Q-fever abortion waves. Further research on historical serological samples from the detected cluster areas, if available, should be performed to confirm occurrence of human Q-fever outbreaks before the May/June 2007 outbreak. In a real-time setting, detection of these clusters should have prompted further investigation and additional laboratory tests, which might have resulted in detection of human Q-fever outbreaks up to 2 years earlier. In this light, it seems worthwhile to make syndromic hospitalization data available real-time for prospective outbreak detection. Finally, our study also illustrates the added value of integrated human and animal surveillance for improved detection of zoonotic disease outbreaks.

NOTE

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org/hyg).

ACKNOWLEDGEMENTS

We thank Dutch Hospital Data for providing data from the Dutch National Medical Register (LMR); Daan Notermans and Marianne van der Sande for reviewing and commenting on the manuscript; and Lucy Phillips for editorial review.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Raoult D, Marrie T, Mege J.** Natural history and pathophysiology of Q fever. *Lancet Infectious Diseases* 2005; **5**: 219–226.
2. **Maurin M, Raoult D.** Q fever. *Clinical Microbiology Reviews* 1999; **12**: 518–553.
3. **Carcopino X, *et al.*** Managing Q fever during pregnancy: the benefits of long-term cotrimoxazole therapy. *Clinical Infectious Diseases* 2007; **45**: 548–555.
4. **van Gageldonk-Lafeber AB, *et al.*** Surveillance of Q fever in the Netherlands [in Dutch]. *Infectieziekten Bulletin* 2003; **14**: 173–177.
5. **Schimmer B, *et al.*** Sustained intensive transmission of Q fever in the south of the Netherlands, 2009. *Eurosurveillance* 2009; **14**: pii=19210.
6. **Karagiannis I, *et al.*** Investigation of a Q fever outbreak in a rural area of The Netherlands. *Epidemiology and Infection* 2009; **137**: 1283–1294.
7. **Schimmer B, *et al.*** Large ongoing Q fever outbreak in the south of The Netherlands, 2008. *Eurosurveillance* 2008; **13**: pii=18939.
8. **Van Steenbergen JE, *et al.*** An outbreak of Q fever in The Netherlands – possible link to goats [in Dutch]. *Nederlands Tijdschrift voor Geneeskunde* 2007; **151**: 1998–2003.
9. **Wouda W, Dercksen DP.** Abortion and stillbirth among dairy goats as a consequence of *Coxiella burnetii* [in Dutch]. *Tijdschrift voor Diergeneeskunde* 2007; **132**: 908–911.
10. **Morroy G, *et al.*** Q-fever: will the mystery be solved? [in Dutch]. *Infectieziekten Bulletin* 2008; **19**: 66–68.
11. **Lazarus R, *et al.*** Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. *BMC Public Health* 2001; **1**: 9.
12. **Heffernan R, *et al.*** Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases* 2004; **10**: 858–864.
13. **van den Wijngaard CC, *et al.*** Syndromic surveillance for local outbreaks of lower-respiratory infections: would it work? *PLoS ONE* 2010; **5**: e10406.
14. **Olson KL, Grannis SJ, Mandl KD.** Privacy protection versus cluster detection in spatial epidemiology. *American Journal of Public Health* 2006; **96**: 2002–2008.
15. **Higgs BW, *et al.*** Early detection of tuberculosis outbreaks among the San Francisco homeless: trade-offs between spatial resolution and temporal scale. *PLoS ONE* 2007; **2**: e1284.
16. **Donker GA.** Continuous morbidity registration sentinels Netherlands 2006 [in Dutch]. NIVEL, Annual Report, 2007.
17. **Rousset E, *et al.*** Coxiella burnetii shedding routes and antibody response after outbreaks of Q fever-induced abortion in dairy goat herds. *Applied and Environmental Microbiology* 2009; **75**: 428–433.
18. **Kulldorff M, *et al.*** A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005; **2**: e59.
19. **Kulldorff M, and Information Management Services Inc.** SaTScanTM v7.0: Software for the spatial and space-time scan statistics (http://www.satscan.org/), 2006. Accessed 18 March 2008.
20. **Abrams AM, Kleinman KP.** A SaTScan macro accessory for cartography (SMAC) package implemented with SAS software. *International Journal of Health Geographics* 2007; **6**: 6.
21. **Kleinman K, Lazarus R, Platt R.** A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology* 2004; **159**: 217–224.
22. **Kulldorff M.** Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society A* 2001; **164**: 61–72.
23. **Gilsdorf A, *et al.*** Large Q fever outbreak due to sheep farming near residential areas, Germany, 2005. *Epidemiology and Infection* 2008; **136**: 1084–1087.
24. **Van den Hoek JA, IJzerman EP, Coutinho RA.** *Legionella* outbreak in Amsterdam: a cooling tower as the source [in Dutch]. *Nederlands Tijdschrift voor Geneeskunde* 2006; **150**: 1808–1811.
25. **La Scola B, Lepidi H, Raoult D.** Pathologic changes during acute Q fever: influence of the route of infection and inoculum size in infected guinea pigs. *Infection & Immunity* 1997; **65**: 2443–2447.
26. **Marrie TJ, *et al.*** Route of infection determines the clinical manifestations of acute Q fever. *Journal of Infectious Diseases* 1996; **173**: 484–487.
27. **Marrie TJ, Pollak PT.** Seroepidemiology of Q fever in Nova Scotia: evidence for age dependent cohorts and geographical distribution. *European Journal of Epidemiology* 1995; **11**: 47–54.
28. **McCaughey C, *et al.*** Human seroprevalence to Coxiella burnetii (Q fever) in Northern Ireland. *Zoonoses Public Health* 2008; **55**: 189–194.

Chapter **6**

Evaluation of Syndromic Surveillance: Perspective on Its Added Value and Recommendations for Implementation

Submitted for publication

Evaluation of Syndromic Surveillance: Perspective on Its Added Value and Recommendations for Implementation.

Cees C. van den Wijngaard[1], Wilfrid van Pelt[1], Nico J.D. Nagelkerke[2], Mirjam Kretzschmar[1], Marion P.G. Koopmans[1,3]

[1]National Institute for Public Health and the Environment, Bilthoven, the Netherlands
[2]United Arab Emirates University, Al-Ain, United Arab Emirates
[3]Erasmus Medical Center, Rotterdam, The Netherlands.

Syndromic surveillance has increasingly been used for capturing infectious disease dynamics that might be missed by surveillance based on notifications by clinicians and laboratory diagnoses alone. There is, however, an ongoing debate about the feasibility of syndromic surveillance and its potential added value. Here we give a perspective on syndromic surveillance, based upon the results of a retrospective analysis of syndromic data from six health-care registries. These registries had been designed for other purposes, but were evaluated for their potential use in signaling 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 independent 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, if feasible in real-time or else retrospectively.

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]. 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) may not alarm clinicians, and uncommon or new pathogens can remain undetected by laboratories (like initially happened with SARS in 2003). Syndromic surveillance may reveal such blind spots of traditional surveillance by monitoring elevations of common symptoms or clinical diagnoses like "shortness of breath" or "pneumonia". The increasing use of syndromic surveillance seems driven by two factors: high profile disease events (2001 anthrax attacks, 2003 SARS outbreak, pandemic threat, excess mortality due to heat waves) stressing the need for an improved early warning surveillance; and the increased availability of electronic health care data, making large scale monitoring of non-specific health indicators increasingly feasible.

There is, however, an ongoing debate about the feasibility and the added value of syndromic surveillance. Some skepticism exists about the potential work load it may generate if used for real-time outbreak detection (i.e. if the system creates many false positive signals) [4]. 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 implementation.

The research questions addressed were:

1. What syndromic data types track known dynamics of infectious diseases in the general population, and thus will also likely reflect emerging pathogen activity?
2. Can syndromic surveillance improve the monitoring of disease burden and/or detect shifts in virulence of common pathogens?
3. Can syndromic surveillance detect (local) outbreaks with a limited number of signals in time, independent of laboratory detection of the causative pathogens?

We addressed these questions by retrospectively analyzing syndromic data from Dutch health-care registries, and also by applying syndromic surveillance to ad-hoc upcoming infectious disease problems. To select potential syndromic data sources, Dutch health-care-registry owners were asked to provide information on predefined criteria (coverage, timeliness, potential for transition to real-time data availability). Table 1 shows the six registries included in the study, with data on work-absenteeism, General Practitioner (GP) consultations, pharmacy prescriptions, laboratory submissions, hospital diagnoses and mortality. Data was available for 1999-2009 or parts of this period.

Based on a literature search, we selected syndromes that were expected to reflect the clinical presentations of both high-threat and common pathogens, and therefore seemed most suitable for syndromic surveillance purposes [5,6]; thus we selected respiratory syndromes (e.g. for high-threat pathogens like *B. anthracis* or pandemic influenza), gastro-enteritis syndromes (e.g. for *Vibrio cholerae)* and neurological syndromes (e.g. for West-Nile virus) for evaluation. The syndromes were defined for each registry guided by a list of syndrome definitions defined by the Centers for Disease Control and Prevention (CDC, USA, http://www.bt.cdc.gov/surveillance/syndromedef) and expert opinion. These registry syndromes were then evaluated for their potential use in signaling infectious disease dynamics and outbreaks.

In this article we give a perspective on the added value of syndromic surveillance for infectious disease surveillance and control, based upon the results of our project and in light of the current literature.

Main findings of syndromic surveillance evaluation

Tracking infectious disease dynamics in the general population

The first question addressed was to what extent trends in respiratory, gastro-enteritis and neurological syndromes in the various registries reflect known pathogen activity, as measured by laboratory pathogen counts. This indicates whether these registries have the potential to reflect emerging pathogen activity (Table 2).

The respiratory syndromes associated best with known disease dynamics (Table 2), displaying higher levels in winter corresponding with higher laboratory counts of respiratory pathogens [7]. Up to 86% of syndrome variations were explained by respiratory pathogen counts, particularly influenza and RSV, which is in line with other studies [8,9]. Our respiratory syndromes were 0-5 weeks ahead of laboratory counts of influenza, suggesting better timeliness of the syndromes. For RSV, the pathogen counts were concurrent with the hospital syndrome, in line with the fact that most RSV tests are performed on hospitalized young children [10,11]. Most other registry syndromes lagged behind the RSV counts, and thus behind the hospital syndrome, which suggests that young children are affected relatively early in the annual RSV season.

Our gastro-enteritis syndromes showed winter peaks concurrent with increased rotavirus activity,

and summer peaks concurrent with peaks in *Shigella*, *Campylobacter* and *Salmonella* activity (Table 2). The syndrome variation 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 syndrome counts 1-2 weeks ahead of the rotavirus laboratory counts [12].

Our general syndromes on infectious neurological disease did not clearly reflect known patterns of pathogen activity, but a more specific viral neurological syndrome did (Table 2). For this "unexplained viral meningitis" syndrome in our hospital data, 62% of the variation was explained by known seasonal enterovirus activity, suggesting that elevated levels of "unexplained viral meningitis" indicate undiagnosed enterovirus infections [13].

Monitoring disease burden and detecting virulence shifts

Relating time series of syndromic surveillance data with pathogen specific surveillance data allows quantifying the disease burden of common pathogens in time. Thus, the monitoring of disease burden and also shifts in virulence can be improved. An example is the clear association of norovirus with mild to severe morbidity and even deaths in the elderly, observed in recent years and coinciding with emergence of new norovirus variants [14]. The latter had been suspected but could not be assessed previously by any other routine surveillance. For influenza, we detected previously unknown shifts in the annual numbers of hospitalizations and deaths related to the number of ILI cases, coinciding with shifts in antigenicity of circulating viruses [15]. Such analyses can also be used for investigating the severity of new influenza A(H1N1) 2009 infection compared with seasonal influenza [16].

Detecting local outbreaks

Obviously, the prime objective of syndromic surveillance is to detect unexpected disease outbreaks in a timely manner. For this purpose, analysis of nationwide data may not be a very sensitive method. Local detection of syndrome elevations — when they are (still) too small to be detected on the national level — might signal emerging outbreaks earlier. To test this, we used known outbreaks of Legionnaires' disease as "positive controls" (or gold standard) for realistic severe respiratory disease outbreaks by uncommon or new pathogens that may not be (timely) detected by traditional surveillance. Simulating prospective surveillance, we were able to timely detect these known gold-standard outbreaks in syndromic hospital data using space-time scan statistics [17]. The fact that the overall alarm rate was modest (on average 5 local clusters detected per year) suggests that a syndromic surveillance on hospitalization data can indeed be a useful early-warning tool for local outbreak detection. By the same approach, previously unknown disease clusters plausibly due to Q fever were detected [18], thus illustrating that in some occasions syndromic surveillance can identify outbreaks that otherwise 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, assessing the absence or (limited) size of unusual disease events can be important as well. For example, Blendon et al. [19] suggested that better communication to the public during the 2003 SARS outbreak might have prevented economic loss due to unnecessary precautions by the public, like staying away from crowds in areas with low level of spread. Also during high-profile public events (like the Olympics, or G8 summit) [20,21], syndromic surveillance should mainly be used to confirm the absence of (major) unusual disease outbreaks. During our project, we illustrated the value of syndromic data for assessing the absence or limited size of unusual disease triggered by ad-hoc upcoming concerns about the possible emergence of specific diseases. For West-Nile virus (WNV), an enhanced surveillance was established by laboratory testing of cerebrospinal fluids (CSFs) from patients with unexplained viral meningitis/encephalitis [13]. None of the CSFs tested positive for WNV, but the probability that WNV was indeed absent in the Netherlands could only be assessed from the annual count of unexplained viral meningitis/encephalitis cases (in relation to the number of CSFs tested). For hepatitis E and Ljungan virus, we evaluated whether considerable hidden viral activity might have occurred, by inspecting time series of unexplained hepatitis and abortion/perinatal death respectively. For impetigo, rumors about a continuing increase in children were countered by inspection of time series of GP consultations for impetigo.

Other spin-offs of syndromic surveillance

In addition to the above described applications, other uses of syndromic surveillance were illustrated during the recent influenza pandemic. We used respiratory syndromic data on hospitalizations 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 [22]. Also early in the pandemic, the reaction of the public to media reports on pandemic influenza was illustrated by sharp elevations of oseltamivir prescriptions [23]. This information was used to urge physicians to exercise restraint in prescribing oseltamivir, in order to decrease the risk of oseltamivir shortage and resistance later in the 2009 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 most for local outbreak detection. Here, relatively small artifacts can already result in false alarms, for example duplicates of the same patient in one registry, as we experienced when using hospital data for space-time cluster detection [17,18]. In a real-time setting (e.g. daily or weekly data updates), reporting delays can also lead to data artifacts and false alarms, if for example hospitals submit their data with a delay [24]. In addition to having few data artifacts, data needs to be representative, and for local outbreak detection also needs 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 [25,26]. Nordin et al. [25] used simulated anthrax attack data injected in true physician's visit data to show that the sensitivity for respiratory outbreaks initiated by bioterrorism was not very high. However, the authors evaluated a maximum system coverage of only 36% of the population. In another study, Balter et al. [26] reported that a syndromic surveillance system in New York City sometimes missed several gastro-enteritis outbreaks due to data quality and coverage problems, such as miscoding of patients' chief complaints and hospitals that did not participate in the system.

For effective signal verification, sufficient information on individual patients' characteristics and concurrent laboratory trends have 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 [17]. Also, the rise in oseltamivir prescriptions early in the 2009 pandemic could be ascribed to the "worried well", because influenza activity had not increased in the laboratory surveillance [23]. Without such verification options, the value of syndromic surveillance is limited [26].

Perspective

Cost effectiveness of real-time systems

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

Estimating the costs of a surveillance system is quite well possible. Studies that report the operating costs associated with real-time syndromic surveillance found annual operating costs ranging from $130.000-$150.000 to $280.000 [27]. However, estimating its benefits is less obvious. Kaufmann et al. [28] reported that the economic damage caused by a bioterrorist attack can amount to millions or billions of dollars. The SARS epidemic in 2003 and the influenza A(H1N1) pandemic in 2009 showed that the damage caused by naturally occurring outbreaks can be similarly high [29,30]. 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 likely 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 for the SARS outbreak or for the 2009 influenza pandemic. Simulation studies could help to further evaluate for what specific types of major disease events syndromic surveillance could likely lead to interventions that limit the damage.

Possibly just as important as benefits by 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 WHO requires, http://www.who.int/csr/disease/influenza/PIPGuidance09.pdf) by reliable syndromic hospital surveillance of severe respiratory infections, costly interventions like quarantine and prophylactic treatment or vaccination could be downscaled or stopped earlier.

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 allowed prospective implementation at relatively low extra costs. Real-time data collection is currently also being implemented for the Dutch GP registry (Table 1). Possibly, including hospital data and other data types in future syndromic surveillance systems is feasible at limited costs as well, if the data collection can be integrated into already planned real-time (future) data infrastructures like the Dutch national health-information-exchange system (EPD, http://www.minvws.nl/dossiers/elektronisch-patienten-dossier).

Conclusion

Based on our evaluation project, we recommend the use of syndromic surveillance to reveal blind spots of traditional surveillance, in particular by detecting unusual (local) outbreaks independent of diagnoses of specific pathogens, and by monitoring of disease burden and virulence shifts of common pathogens.

Our results are mostly based on retrospective analysis of syndromic data with high quality and coverage. If prospective collection of syndromic data of high quality and coverage is not feasible, real-time early warning for (local) outbreaks should not be performed, since true outbreaks will then likely 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 for syndromic signals, and thus recognize relevant signals that might need a response. Retrospective analyses as performed in our project can validate the involved data and analyses before prospective implementation.

Acknowledgments

We thank Statistics Netherlands (CBS), 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.

Table 1. Syndromic data registries included in the syndromic surveillance evaluation study.

Data type	**Period**	**Coverage** (16.3 million pop)	**Syndrome definitions** (detailed syndrome definitions and codes available on request)	**Analyzed data**	**International code system**	**Registry**	**Prospective implementation**
Absenteeism	2002-2003	80% (of working population, 8 million)	Employers reported sick, no further medical information	Sick leave reports of employers	-	Statistics Netherlands (CBS) (http://www.cbs.nl)	n/a*
General Practice consultations	2001-2004	1-2%**	Symptoms and diagnoses indicating infectious disease	Symptoms and diagnoses recorded in practice or telephone consultations, and home visits	ICPC (International Classification of Primary Care)	Netherlands Information Network of General Practice (LINH) (http://www.nivel.nl/linh/)	Real-time system currently being implemented
Pharmacy dispensations	2001-2003	85%	Prescribed medications indicative for infectious disease	Prescription medications dispensed in Dutch pharmacies, coded according to the WHO Anatomical Therapeutic Chemical (ATC) classification	ATC (Anatomical Therapeutic Chemical Classification System)	Foundation for Pharmaceutical Statistics (http://www.sfk.nl)	Currently monthly data updates are feasible in ad-hoc public health situations.
Hospitalizations	1999-2007	99%	General symptoms/diagnoses(cdc1) *and* specific biologic agent diagnoses (cdc3)	Discharge and secondary diagnoses and date of hospitalization.	ICD-9-CM (International Classification of Diseases, 9th revision, Clinical Modification)	Dutch National Medical Register (LMR, http://www.dutchhospitaldata.nl/)	No prospective implementation possible in short term (annual data updates will continue)
Laboratory submissions (negative and positive results)	2001-2004	16%	Submissions for microbiological diagnostic tests on materials; *and* all submissions for serology on known specific pathogens.	Lab.submission requests for diagnostic testing	-	National Infectious Diseases Information System (ISIS)	n/a*
Mortality	1999-2004; crude (overall) mortality 1999-2009	100%	General symptoms/diagnoses *and* specific biologic agent diagnoses	Date of death, primary cause of death, complicating and other additional causes of death	ICD-10 (Internat. Classification of Diseases, 10th revision)	Statistics Netherlands (CBS) (http://www.cbs.nl)	Currently weekly crude mortality surveillance, pilot phase.

*The Lab-requests registry (ISIS-labs) and the absenteeism registry stopped to exist during the project.
**The GP registry coverage will become 5% in the next few years as part of the Flu-pandemic project.

Table 2. Tracking infectious disease dynamics in the general population of the Netherlands: respiratory, gastro-enteritis and neurological syndromes in the six included health-care registries. The table describes for each type of data and syndrome 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.*

	Syndromes		
Registries	**general respiratory syndrome**	**general gastro-enteritis syndrome**	**general neurological syndrome**
Absenteeism	winter peaks concurrent with peaks in influenza, RSV and other respiratory pathogens; 68% of variations explained by respiratory pathogens; 2 weeks ahead of RSV, 4-5 weeks ahead of influenza [7]	not evaluated	not evaluated
General Practice consultations	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 [7]	winter peaks and summer peaks, concurrent with rotavirus and shigella/salmonella/campylobacter peaks respectively; 29% of variations explained by gastro-enteral pathogens (51% for the 0-4 years of age, 2 weeks ahead of rotavirus) [12]; an increase in winter 2002/03 possibly related to norovirus activity [14]	no clear reflection of known disease dynamics
Pharmacy dispensations	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 [7]	relatively low winter peaks and higher summer peaks, concurrent with rotavirus and shigella/salmonella/campylobacter peaks respectively	not evaluated
Hospitalizations	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 ahead of influenza [7]	relatively high winter peaks and lower summer peaks, concurrent with rotavirus and shigella/salmonella/campylobacter peaks respectively; 40% of variations explained by gastro-enteral pathogens (85% for the 0-4 years of age, 1 week ahead of rotavirus) [12]; an increase in winter 2002/03 possibly related to norovirus activity [14]	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 the enterovirus notifications [13]
Laboratory submissions (negative and positive results)	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 weeks ahead of influenza [7]	relatively low winter peaks and higher summer peaks, concurrent with rotavirus and shigella/salmonella/campylobacter peaks respectively	no clear reflection of known disease dynamics
Mortality	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 weeks ahead of influenza [7]	no obvious reflection of known (seasonal) pathogen activity; an increase in winter 2002/03 possibly related to norovirus activity [14]	no clear reflection of known disease dynamics

*The timeliness differences were assessed by the optimized lags of pathogen counts in time series models that explain the syndrome variation [7].

Reference List

1. Cooper DL, Verlander NQ, Smith GE, Charlett A, Gerard E, Willocks L et al. Can syndromic surveillance data detect local outbreaks of communicable disease? A model using a historical cryptosporidiosis outbreak. Epidemiol Infect 2006;134:13-20.
2. 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.
3. Bellika JG, Hasvold T, Hartvigsena G. Propagation of program control: A tool for distributed disease surveillance. Int J Med Inform 2007;76:313-29.
4. Hope K, Durrheim DN, d'Espaignet ET, Dalton C. Syndromic Surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 2006;60:374-5.
5. 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.
6. Wagner MM, Dato V, Dowling JN, Allswede M. Representative threats for research in public health surveillance. J Biomed Inform 2003;36:177-88.
7. Van den Wijngaard C, van Asten L, van Pelt W, Nagelkerke NJ, Verheij R, de Neeling AJ et al. Validation of syndromic surveillance for respiratory pathogen activity. Emerg Infect Dis 2008;14:917-25.
8. Bourgeois FT, Olson KL, Brownstein JS, McAdam AJ, Mandl KD. Validation of syndromic surveillance for respiratory infections. Ann Emerg Med 2006;47:265. e1.
9. 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.
10. Terletskaia-Ladwig E, Enders G, Schalasta G, Enders M. Defining the timing of respiratory syncytial virus (RSV) outbreaks: an epidemiological study. BMC Infect Dis 2005;5:20.
11. Simoes EA. Respiratory syncytial virus infection. Lancet 1999;354:847-52.
12. Van Asten L, van den Wijngaard C, van Pelt W, van Vliet H, Koopmans M. Understanding the dynamics of gastro-intestinal syndrome: general practioner and hospital data vs laboratory surveillance. Advances in Disease Surveillance 2007;2:215. Available at http://www.isdsjournal.org/article/viewFile/949/615, accessed at May 28th 2010.
13. Rockx B, Asten LV, Wijngaard CV, Godeke GJ, Goehring L, Vennema H et al. Syndromic surveillance in the Netherlands for the early detection of west nile virus epidemics. Vector Borne Zoonotic Dis 2006;6:161-9.
14. Van Asten L, van den Wijngaard C, Siebenga J, van Pelt W, van Vliet H, Koopmans M. Greater pathogenicity of norovirus strains in 2003? A syndromic approach. Advances in Disease Surveillance 2007;2:175. Available at http://www.isdsjournal.org/article/view/2825/2143, accessed at May 28th 2010.
15. Van den Wijngaard CC, van Asten L, Meijer A, van Pelt W, Nagelkerke NJD, Donker GA et al. Detection of Increased Severity of Influenza Infections: Associating Respiratory Hospitalizations and Mortality with Influenza-Like-Illness in Primary Care. Am J Public Health 2010; (in press).
16. Van 't Klooster TM, Wielders CC, Donker T, Isken L, Meijer A, van den Wijngaard CC et al. Surveillance of hospitalisations for 2009 pandemic influenza A(H1N1) in the Netherlands, 5 June - 31 December 2009. Euro Surveill 2010;15:pii=19461.
17. Van den Wijngaard CC, van Asten L, van Pelt W, Doornbos G, Nagelkerke NJ, Donker G et al. Syndromic surveillance for local outbreaks of lower-respiratory infections: Would it work? PLoS ONE 2010;5:e10406.
18. Van den Wijngaard CC, Dijkstra F, van Pelt W, van Asten L, Kretzschmar M, Schimmer B et al. In search of hidden Q-fever outbreaks: linking syndromic hospital clusters to infected goat farms . Epidemiol Infect 2010; (epub).
19. Blendon RJ, Benson JM, DesRoches CM, Raleigh E, Taylor-Clark K. The public's response to severe acute respiratory syndrome in Toronto and the United States. Clin Infect Dis 2004;38:925-31.
20. Gesteland PH , Gardner RM, Tsui FC, Espino JU, Rolfs RT, James BC et al. Automated Syndromic Surveillance for the 2002 Winter Olympics. J Am Med Inform Assoc 2003;10:547-554.
21. Meyer N, McMenamin J, Robertson C, Donaghy M, Allardice G, Cooper D. A multi-data source surveillance system to detect a bioterrorism attack during the G8 Summit in Scotland. Epidemiol Infect 2008;136:876-85.
22. Van Asten L, van der Lubben M, van den Wijngaard C, van Pelt W, Verheij R, Jacobi A et al. Strengthening the diagnostic capacity to detect Bio Safety Level 3 organisms in unusual respiratory viral outbreaks. J Clin Virol 2009;45:185-90.
23. Van den Wijngaard CC, van Steenbergen JE, van der Sande MA, Koopmans MP. [New influenza A (H1N1): advised indication and prescription of antiviral drugs]. Ned Tijdschr Geneeskd 2009;153:A1053.
24. Travers DA, Barnett C, Ising A, Waller A. Timeliness of emergency department diagnoses for syndromic surveillance. AMIA Annu Symp Proc 2006;769-73.
25. Nordin JD, Goodman MJ, Kulldorff M, Ritzwoller DP, Abrams AM, Kleinman K et al. Simulated anthrax attacks and syndromic surveillance. Emerg Infect Dis 2005;11:1394-8.
26. Balter S, Weiss D, Hanson H, Reddy V, Das D, Heffernan R. Three years of emergency department gastrointestinal syndromic surveillance in New York City: what have we found? MMWR Morb Mortal Wkly Rep 2005;54 Suppl:175-80.
27. Kirkwood A, Guenther E, Fleischauer AT, Gunn J, Hutwagner L, Barry MA. Direct Cost Associated With the Development and Implementation of a Local Syndromic Surveillance System. J Public Health Manag Pract 2007;13:194-9.
28. Kaufmann AF, Meltzer MI, Schmid GP. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? Emerg Infect Dis 1997;3:83-94.
29. Dan YY, Tambyah PA, Sim J, Lim J, Hsu LY, Chow WL et al. Cost-effectiveness analysis of hospital infection control response to an epidemic respiratory virus threat. Emerg Infect Dis 2009;15:1909-16.
30. Brouwers L, Cakici B, Camitz M, Tegnell A, Boman M. Economic consequences to society of pandemic H1N1 influenza 2009 - preliminary results for Sweden. Euro Surveill 2009;14:pii= 19333.

Chapter 7

General Discussion

General Discussion

Introduction

Syndromic surveillance has increasingly been used for capturing infectious disease dynamics [1-8] and the impact of other public-health events, like heat waves or chemical incidents [7,9,10]. In chapters 2, 3, 4 and 5 of this thesis the feasibility of several syndromic surveillance applications was investigated and in chapter 6 we concluded that syndromic surveillance — either in real-time or retrospectively — can be used for 1) monitoring disease dynamics in the general population, 2) monitoring disease burden and/or virulence shifts and 3) detection of (local) outbreaks. Below we will further elaborate on the implications of this thesis for syndromic surveillance regarding *data, analyses and response*, in light of the current literature. Then we will discuss challenges and future possibilities of syndromic surveillance, and finally will conclude upon the added value of syndromic surveillance for infectious disease surveillance.

Data

The included data types and quality, as well as the syndrome definitions can greatly impact the performance of syndromic surveillance systems. The syndromic data types investigated in this thesis all more or less reflected infectious disease dynamics in the general population. This has also been observed for other health-care data types, like over-the-counter medication sales data [11], telephone help-line data [12,13], and hospital emergency-department data [3,4,14]. A recent trend is the use of internet-based data sources, like search terms [15-17], media reports [18], or web-based questionnaires [19-21]. Internet-based data types capture disease dynamics in the community more or less independent of health-care-seeking behavior of the population.

An important difference between the various syndromic data types is that some reflect mild morbidity, not even requiring medical treatment, whereas others reflect more severe morbidity or even mortality. Besides a difference in severity level, this also implies differences in vulnerability to biases like reactions by the "worried well" due to media attention for major infectious disease events, as discussed in chapter 6. For this reason, signal verification is especially important for mild morbidity data, whereas for internet-based data this is a weak point; for searches and media reports signal verification on patient level is difficult. Also, web-based applications often provide biased datasets due to lack of coverage in the very young and the elderly that do not use internet routinely [19,20]. Finally, for internet searches or media reports, standardized metrics such as disease incidence and prevalence do not apply, although they are being developed [22]. For mild morbidity data based on health-care registries (GP consultations, medical prescriptions) signal verification on patient level is less of a problem, since epidemiological data on patient level (age, place of residence, specific symptoms etc.) will be available, and additional laboratory testing is often possible.

The sensitivity and specificity of (local) outbreak detection are also strongly influenced by the included data types. Using milder morbidity data may lead to low specificity; numerous alarms will be generated that only reflect space-time variations in common mild disease that do not necessarily need a public-health intervention. Thus, a way to increase the specificity is to focus on data reflecting severe morbidity (like hospitalizations or mortality) hereby limiting the alarm rate. Then again, that may decrease the sensitivity, since data on severe morbidity will miss early signals generated by local increases of still mild but progressing disease. Also, for

monitoring the disease burden or shifts in virulence of common pathogens by combined analysis of syndromic data and specific (pathogen) surveillance (chapter 3 and [23]), data ranging from mild to severe morbidity should be available, to be able to assess the impact of pathogens at all morbidity levels.

Data quality is most important for local outbreak detection, since small artifacts can already result in false alarms (see data requirements described in chapter 6). Besides quality parameters like few data artifacts and high coverage, a higher spatial resolution in the data reportedly improves outbreak detection, but there is a trade-off with privacy concerns [24,25].

We focused on respiratory, gastro-enteritis and neurological syndrome definitions, since these syndromes represent clinical presentations of both high-threat and common pathogens (chapter 1). Nevertheless, for all syndromic surveillance applications and data types, syndrome definitions should be flexible, to be able to adapt to new unforeseen disease events. For example, in our hospitalization data we could easily adapt the general neurological syndrome definition based on CDC-guidelines [26] into a more specific viral meningitis/encephalitis syndrome, which made it possible to assess the validity of an enhanced West-Nile virus surveillance [27].

Syndromic data can be available as structured (coded) data or free-text data. Surveillance on structured data has shown good performance in our and other studies (chapter 2 and [28-30]); syndrome definitions are often based upon combinations of structured registrations of symptoms or (clinical) diagnoses (like ICD-9-codes). Surveillance based on free-text data [28,31-33] is methodologically more challenging since it requires natural language processing to encode the data text fields, and can lead to mixed results. For instance, Hripcsak et al. found that free-text syndromes based on electronic health records correlated well with influenza-like-illness, but not with gastro-intestinal infections [28]. Nevertheless, if structured data is lacking, using free-text data makes it possible to implement syndromic surveillance as long as the accuracy of syndrome classifiers is properly validated.

Analyses

Many time-series methods are being used for syndromic time trends, mostly focusing on aberration detection in time [34]. In our project, rather than aberration detection, we first of all wanted to evaluate to what extent our syndromes reflected known disease dynamics. Therefore, we constructed multivariate regression models that explained the syndrome time series by specific disease or pathogen surveillance data (chapter 2 and 3). When modeling the syndromes we did not transform them to the log scale, to be able to estimate the additive contributions of individual pathogens on the syndromes. In chapter 3, in order to estimate better confidence intervals when using such additive models to search for shifts in disease burden or virulence of common pathogens, we refined our time-series method by using a poisson (instead of linear) error distribution, and including autocorrelation in the model residuals.

For local outbreak detection, we used the space-time permutation and the poisson-distributed scan statistic as available in the SaTScan software (chapters 4 and 5). Although these scan-statistics are currently among the most widely used methods for space-time surveillance systems, they have a disadvantage that they can only search for circular- (or elliptic-) shaped clusters, which makes them less sensitive to clusters that are shaped otherwise. That is why Takahashi et al. proposed an alternative scan statistic for irregularly shaped clusters [35]; however, their method is not practical to apply on large data sets, because of its long computation time. Recently an alternative scan-statistic approach has been described by Neill and Cooper that might solve

both the problem of irregularly shaped clusters and the problem of long computation time, but this method has not been evaluated yet in real-life surveillance systems [36].

Further research is needed in methodology for analyzing multiple data streams [36-41]. Many studies on syndromic surveillance analyze disease dynamics in univariate syndromic time series, whereas multiple data streams are often available. Shifts in single syndromes can easily be biased by shifts in health-care utilization (e.g. during major public events), whereas the relation between multiple syndromes will likely be more stable during shifts in health-care utilization. Multivariate analyzes can therefore provide better information on disease dynamics; for instance, by focusing on shifts in the relation between multiple syndromes, instead of focusing on mean shifts in single syndromes, thus raising the specificity of signals [37,40]. Similarly, multivariate methods exist for space-time analysis, to detect outbreaks in multiple data streams at the same time [36,38].

Response

If syndromic surveillance is used for early warning, responding on signals can be complicated, because the cause and thus possible threat of a syndromic signal initially is often unknown. No standardized response protocol for syndromic surveillance signals is available in the current literature [42,43]. A logical stepwise approach, after excluding obvious data artifacts, would be:

1. to verify whether explanations can be found in readily available laboratory and epidemiological data (like patients' demographics and clinical diagnoses) or other relevant data sources (e.g. climate data for possibly temperature related morbidity/mortality)
2. to contact the patients' physicians to ask for possible explanations for the signal
3. to perform additional laboratory diagnostics to identify possible causative pathogens or other causes.

Step 1, verification using epidemiological and laboratory (and other relevant) data, would often be sufficient to decide that no further investigation is necessary. For example, for local respiratory hospital signals, we described that assessing activity of RSV or influenza overlapping with syndromic signals would often make further investigation unnecessary (see chapter 4). In step 2, identification of patients or their physicians is necessary. Whether this is allowed will depend on national privacy regulations which may differ between countries. For example in Sweden, personal identity numbers and even patients' addresses can be used for surveillance purposes [44], whereas in many other countries this would not be allowed. Step 3, performing additional diagnostics, would only be necessary if no explanation has yet been found, but might be hard if patients are not in the immediate surrounding of health-care facilities (e.g. hospitalized). Cooper et al. describe a possible solution for this; they successfully obtained additional laboratory results by sending self-sampling test kits to "cold/flu" patients who called in to a telephone health-help line [45].

When using real-time total mortality monitoring as an early-warning tool for excess mortality during the 2009 pandemic, we experienced that we could not follow the above described response steps, mainly due to privacy regulations. We did observe some small but persistent elevations in infant mortality (less than one year of age, and scattered all over the country), for which we wanted to verify if they might be related to the pandemic [46]. Privacy regulations did not allow us to verify any further detailed epidemiological, clinical, or laboratory data for these patients (step 1). Without such basic information, it may have been difficult to convince the reporting physicians (step 2) to supply additional information for verification. More

importantly, identifying the reporting physicians was not allowed by privacy regulations. Obviously, without identification of patients or their physicians, additional diagnostics (step 3) were not possible either. Because the cause of this signal remained uncertain, no specific further action was taken.

This example illustrates that small to moderate syndromic signals that cannot be verified in epidemiogical or laboratory data, or by consulting patients' physicians, will have to be ignored by public-health authorities, as without such information irrelevant signals cannot be distinguished from relevant signals. Nevertheless, syndromic surveillance systems with similar privacy or practical limitations as described above can still be helpful for public-health decision making; our real-time surveillance on mortality data did support the decision making during the 2009 pandemic, particularly when weighing the (limited) severity of disease against the impact of control measures. Suppose that the observed excess mortality would have been higher and/or more persistent, then additional response measures could have been imposed — even without detailed signal verification — such as increased and sustained active case finding and prophylactic treatment of contacts. For each system to be implemented it should be investigated how surveillance can best be performed within the privacy regulations involved, or whether the surveillance purposes are relevant enough for public health to maybe adapt the privacy regulations.

In a prospective setting, responses will be based on a combination of syndromic surveillance data, reports by clinicians and early diagnostic test results, which all could generate the first alarm for different kind of disease events. Further research should focus on practical methods for integrating syndromic surveillance alarms — preferably from systems with high data quality and coverage as in our retrospective study — with clinical reports and laboratory information for effective public-health responses.

Performance evaluation

One of the major concerns is that syndromic surveillance could have low specificity resulting in many false alarms [47-50]. As described in the introduction (chapter 1), standardizing performance evaluation is difficult, also because real-life test outbreaks are scarce, while simulated outbreaks do not reflect real-life diversity and unpredictability. Therefore, quantitative performance measures like sensitivity and specificity of detection cannot be generalized. In our project we chose to evaluate local outbreak detection using the few recent known outbreaks of severe respiratory infectious diseases in the Netherlands. We then formally assessed whether these true outbreaks could be detected (as an indication for sensitivity), and calculated the average overall alarm rate in time (as an indication for specificity, or the work load of investigating these alarms, see chapter 4). Buckeridge et al. attempted to improve performance evaluation for temporal aberration detection algorithms; they designed a conceptual model that decomposes the different tasks of the algorithms, to facilitate the identification of the precise characteristics that determine their performance (e.g. whether forecasting is used to generate an expected value, the forecasting method etc.) [34]. They implemented this model into software to conduct detection performance evaluation studies in a formal systematic way, which might lead to a more standardized evaluation of detection algorithms in the future.

Buehler et al. interviewed US syndromic surveillance users to assess qualitatively for what kind of disease events syndromic surveillance systems were considered useful in public-health practice [51]. These users reported that they valued syndromic surveillance systems most for monitoring widespread health effects and affirming the absence of outbreaks in crisis situations ("situational awareness"). This is in line with

our own experiences (see chapter 6), and suggests that, besides a system's outbreak detection performance, the ability to quickly address ad-hoc concerns should also be weighed (e.g. by flexible syndrome definitions that can be real-time adapted to focus on specific diseases).

Integrated surveillance

In the future, instead of implementing systems with only syndromic or only laboratory data, it would make sense to use integrated surveillance data from different sources for better situational awareness and early warning. In this light, one could for example think of first detecting the emergence of a new variant of a specific pathogen, and then monitoring its clinical impact by integrated surveillance of both molecular and syndromic data. Some studies already presented ideas in this direction. Rabadan et al. suggested that the increased availability of patient data from Electronic Health Records together with increased availability of molecular typing techniques could facilitate an integrated pandemic influenza surveillance to evaluate and understand the severity of the pandemic virus and to identify the populations at risk of mild or severe, life-threatening illness [52]. Retrospective analyses of the 2009 pandemic could show to what extent this approach results in additional information about its impact. Ansaldi et al. showed that syndromic surveillance combined with molecular typing can improve the early-warning and tracing back the origin of measles outbreaks [53].

Several studies suggested that syndromic *animal* surveillance can improve the preparedness for human disease events ([27,54-56] chapter 5). Gubernot et al. presented a framework for an integrated animal-human disease surveillance in the US, for a specific list of zoonotic agents (such as *B. anthracis*, *C. botulinum* toxin, *Brucella spp.*) [57]. This framework was presented as a plan that only becomes active during times of increased threat for specific agents, with temporary extra work-load for the involved public-health, health-care and veterinary staff. Implementing a useful *routine* integrated human-animal surveillance system with an acceptable extra work-load seems more challenging. In the Netherlands, syndromic animal surveillance on poultry farms has already been implemented after the influenza-A(H7N7) outbreak in 2003 (mainly based on mandatory reporting of increased mortality) [58]. For better preparedness to future threats of avian influenza, integration of such syndromic animal and human surveillance systems seems worthwhile, since detection of human severe respiratory infection clusters in proximity of poultry farms with increased mortality could trigger earlier laboratory confirmation. For an effective response to signals from such integrated human-animal surveillance, an extra challenge could be to overcome jurisdictional barriers and conflicts of interest between human and veterinary public health.

Feedback to local health-care staff

Besides improving the situational awareness and early warning surveillance for national or regional public-health authorities, syndromic surveillance could fulfill the same purpose for local hospital and ambulatory-care staff, or even the local public. Johansen et al. reported that patients in Norway expected their GP to know the current disease dynamics in the local population, which might be relevant for their consult, whereas the GP's in this study stated that they actually were not systematically informed of local disease dynamics [59]. Horst et al. suggested tracking the spread of common illnesses by syndromic surveillance and GIS to improve local health-system resource allocation and inform the public [60]. A positive side

effect of applying syndromic surveillance tools for local purposes could be that health-care staff will be extra motivated to improve the quality of data that goes into the system (e.g. electronic health records) if they find the output useful for their own daily practice [61]. It should thus be further investigated whether health-care workers could indeed use local syndromic surveillance tools for their daily practice in a useful manner.

Conclusion

In the last decade, syndromic surveillance systems worldwide have improved the situational awareness for (infectious) disease dynamics and outbreaks. The results of our analyses confirm that syndromic data can reflect disease dynamics that remain undetected by traditional surveillance alone, and also indicate that early outbreak detection is feasible if syndromic data with high quality and coverage can be collected. For an effective signal verification and response, preferably timely epidemiological and laboratory information should be available, and it should be possible to contact patients or their physicians as well.

If syndromic data is not real-time available yet, retrospective data collection and analysis can still be of added value for emerging infectious disease control. Emerging diseases can become persistent or recurrent in affected populations, especially if their introduction is initially not detected. Our results showed that the Q-fever outbreak that was first detected in 2007, could possibly have been detected in syndromic hospital data from 2005. Even with the current lag time before this hospital data becomes available (once a year), these clusters could already have been detected in 2006, which is still one year before the official recognition of the outbreak. For pandemic influenza A(H1N1), retrospective syndromic analyses could be used for prospective public-health responses; by doing a comparative analysis of the pandemic impact in 2009 compared to seasonal influenza in previous years, it becomes possible to reconsider interventions measures like additional vaccination or treatment if the same pandemic influenza strain would return in 2010/11. Public-health authorities should therefore periodically collect and analyze available retrospective syndromic data from mild to severe morbidity as long as no real-time substitute is available.

In this thesis, we illustrated how syndromic data can be used complementary to laboratory data to reveal infectious disease dynamics and outbreaks that otherwise would remain undetected. Syndromic data in currently existing real-time systems often lacks coverage and quality. For future inclusion of better syndromic data for surveillance, further research is necessary to assess to what extent the growing development of health–information-exchange and electronic-health-records might facilitate data collection at limited costs, with no or minimal extra work-load for health-care personnel, but without crossing legal or ethical boundaries regarding privacy.

Reference List

1. Muscatello DJ, Churches T, Kaldor J, Zheng W, Chiu C, Correll P, et al. An automated, broad-based, near real-time public health surveillance system using presentations to hospital Emergency Departments in New South Wales, Australia. BMC Public Health 2005;5:141.
2. Cooper DL, Verlander NQ, Smith GE, Charlett A, Gerard E, Willocks L, et al. Can syndromic surveillance data detect local outbreaks of communicable disease? A model using a historical cryptosporidiosis outbreak. Epidemiol Infect 2006;134:13-20.
3. 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.
4. Van-Dijk A, Aramini J, Edge G, Moore KM. Real-time surveillance for respiratory disease outbreaks, Ontario, Canada. Emerg Infect Dis 2009;15:799-801.
5. Bellika JG, Hasvold T, Hartvigsena G. Propagation of program control: A tool for distributed disease surveillance. Int J Med Inform 2007;76:313-29.
6. Fleming DM, Barley MA, Chapman RS. Surveillance of the bioterrorist threat: a primary care response. Commun Dis Public Health 2004;7:68-72.
7. Flamand C, Larrieu S, Couvy F, Jouves B, Josseran L, Filleul L. Validation of a syndromic surveillance system using a general practitioner house calls network, Bordeaux, France. Euro Surveill 2008;13:pii=18905.
8. Wu TS, Shih FY, Yen MY, Wu JS, Lu SW, Chang KC, et al. Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. BMC Public Health 2008;8:18.
9. Doroshenko A, Cooper D, Smith G, Gerard E, Chinemana F, Verlander N, et al. Evaluation of syndromic surveillance based on National Health Service Direct derived data--England and Wales. MMWR Morb Mortal Wkly Rep 2005;54 Suppl:117-22.
10. Leonardi GS, Hajat S, Kovats RS, Smith GE, Cooper D, Gerard E. Syndromic surveillance use to detect the early effects of heat-waves: an analysis of NHS direct data in England. Soz Praventivmed 2006;51:194-201.
11. 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.
12. Caudle JM, van Dijk A, Rolland E, Moore KM. Telehealth Ontario detection of gastrointestinal illness outbreaks. Can J Public Health 2009;100:253-7.
13. 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.
14. Bourgeois FT, Olson KL, Brownstein JS, McAdam AJ, Mandl KD. Validation of syndromic surveillance for respiratory infections. Ann Emerg Med 2006;47:265.e1.
15. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009;49:1557-64.
16. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457:1012-4.
17. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. PLoS ONE 2009;4:e4378.
18. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009;360:2153-2157
19. Sugiura H, Ohkusa Y, Akahane M, Sugahara T, Okabe N, Imamura T. Construction of syndromic surveillance using a web-based daily questionnaire for health and its application at the G8 Hokkaido Toyako Summit meeting. Epidemiol Infect 2010;138:1493-502
20. Friesema IH, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, et al. Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 2009;27:6353-7.
21. Wethington H, Bartlett P. The RUsick2 Foodborne Disease Forum for syndromic surveillance. Emerg Infect Dis 2004;10:401-5.
22. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11:e11.

23. Van Asten L, van den Wijngaard C, Siebenga J, van Pelt W, van Vliet H, Koopmans M. Greater pathogenicity of norovirus strains in 2003? A syndromic approach. Advances in Disease Surveillance 2007;2:175. Available at http://www.isdsjournal.org/article/view/2825/21432, accessed at May 28th 2010.
24. Olson KL, Grannis SJ, Mandl KD. Privacy Protection Versus Cluster Detection in Spatial Epidemiology. Am J Public Health 2006;96:2002-8.
25. Higgs BW, Mohtashemi M, Grinsdale J, Kawamura LM. Early detection of tuberculosis outbreaks among the San Francisco homeless: trade-offs between spatial resolution and temporal scale. PLoS ONE 2007;2:e1284.
26. Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef/word/syndromedefinitions.doc, accessed at May 28th 2010.
27. Rockx B, van Asten L, van den Wijngaard C, Godeke GJ, Goehring L, Vennema H, et al. Syndromic surveillance in the Netherlands for the early detection of west nile virus epidemics. Vector Borne Zoonotic Dis 2006;6:161-9.
28. Hripcsak G, Soulakis ND, Li L, Morrison FP, Lai AM, Friedman C, et al. Syndromic surveillance using ambulatory electronic health records. J Am Med Inform Assoc 2009;16:354-61.
29. Espino JU, Wagner MM. Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. Proc AMIA Symp 2001;164-8.
30. Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenzalike illness by International Classification of Diseases, Ninth Revision. Emerg Infect Dis 2007;13:207-16.
31. Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. Artif Intell Med 2005;33:31-40.
32. Chapman WW, Dowling JN, Wagner MM. Classification of emergency department chief complaints into 7 syndromes: a retrospective analysis of 527,228 patients. Ann Emerg Med 2005;46:445-55.
33. South BR, Gundlapalli AV, Phansalkar S, Shen S, Delisle S, Perl T, et al. Automated detection of GI syndrome using structured and non-structured data from the VA EMR. Advances in Disease Surveillance 2007;4:62. Available at http://www.isdsjournal.org/article/viewFile/1974/1529, accessed at May 28th 2010.
34. Buckeridge DL, Okhmatovskaia A, Tu S, O'Connor M, Nyulas C, Musen MA. Understanding detection performance in public health surveillance: Modeling aberrancy-detection algorithms. J Am Med Inform Assoc 2008;15:760-9.
35. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. Int J Health Geogr 2008;7:14.
36. Neill DB, Cooper GF. A multivariate Bayesian scan statistic for early event detection and characterization. Mach Learn 2010;79:261-82.
37. Reis BY, Kohane IS, Mandl KD. An Epidemiological Network Model for Disease Outbreak Detection. PLoS Med 2007;4:e210.
38. Kulldorff M, Mostashari F, Duczmal L, Katherine Yih W, Kleinman K, Platt R. Multivariate scan statistics for disease surveillance. Stat Med 2007;26:1824-33.
39. Rolka H, Burkom H, Cooper GF, Kulldorff M, Madigan D, Wong WK. Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs. Stat Med 2007;26:1834-56.
40. Mohtashemi M, Kleinman K, Yih WK. Multi-syndrome analysis of time series using PCA: A new concept for outbreak investigation. Stat Med 2007;26:5203-24.
41. Wang L, Ramoni MF, Mandl KD, Sebastiani P. Factors affecting automated syndromic surveillance. Artif Intell Med 2005;34:269-78.
42. Uscher-Pines L, Farrell CL, Babin SM, Cattani J, Gaydos CA, Hsieh YH, et al. Framework for the development of response protocols for public health syndromic surveillance systems: case studies of 8 US states. Disaster Med Public Health Prep 2009;3:S29-36.
43. Uscher-Pines L, Farrell CL, Cattani J, Hsieh YH, Moskal MD, Babin SM, et al. A survey of usage protocols of syndromic surveillance systems by state public health departments in the United States. J Public Health Manag Pract 2009;15:432-8.
44. Sartorius B, Jacobsen H, Torner A, Giesecke J. Description of a New all Cause Mortality Surveillance System in Sweden as a Warning System Using Threshold Detection Algorithms. Eur J Epidemiol 2006;21:181-9.
45. 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.

46. Mazick A, Gergonne B, Wuillaume F, Danis K, Vantarakis A, Uphoff H, et al. Higher all-cause mortality in children during autumn 2009 compared with the three previous years: pooled results from eight European countries. Euro Surveill 2010;15:pii=19480.
47. Balter S, Weiss D, Hanson H, Reddy V, Das D, Heffernan R. Three years of emergency department gastrointestinal syndromic surveillance in New York City: what have we found? MMWR Morb Mortal Wkly Rep 2005 ;54 Suppl:175-80.
48. Hope K, Durrheim DN, d'Espaignet ET, Dalton C. Syndromic Surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 2006;60:374-5.
49. Buckeridge DL, Owens DK, Switzer P, Frank J, Musen MA. Evaluating detection of an inhalational anthrax outbreak. Emerg Infect Dis 2006;12:1942-9.
50. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterror 2003;1:77-81.
51. Buehler JW, Whitney EA, Smith D, Prietula MJ, Stanton SH, Isakov AP. Situational uses of syndromic surveillance. Biosecur Bioterror 2009;7:165-77.
52. Rabadan R, Calman Md N, Hripcsak G. Next Generation Syndromic Surveillance: Molecular Epidemiology, Electronic Health Records and the Pandemic Influenza A (H1N1) Virus. PLoS Curr Influenza 2009;RRN1012.
53. Ansaldi F, Orsi A, Altomonte F, Bertone G, Parodi V, Carloni R, et al. Syndrome surveillance and molecular epidemiology for early detection and tracing of an outbreak of measles in Liguria, Italy. J Med Virol 2009;81:1807-13.
54. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg Infect Dis 2003;9:641-6.
55. Glickman LT, Moore GE, Glickman NW, Caldanaro RJ, Aucoin D, Lewis HB. Purdue university-banfield national companion animal surveillance program for emerging and zoonotic diseases. Vector Borne Zoonotic Dis 2006;6:14-23.
56. Rabinowitz P, Gordon Z, Chudnov D, Wilcox M, Odofin L, Liu A, et al. Animals as sentinels of bioterrorism agents. Emerg Infect Dis 2006;12:647-52.
57. Gubernot DM, Boyer BL, Moses MS. Animals as early detectors of bioevents: veterinary tools and a framework for animal-human integrated zoonotic disease surveillance. Public Health Rep 2008;123:300-15.
58. Elbers AR, Holtslag JB, Bouma A, Koch G. Within-flock mortality during the high-pathogenicity avian influenza (H7N7) epidemic in The Netherlands in 2003: implications for an early detection system. Avian Dis 2007;51:304-8.
59. Johansen MA, Johnsen JA, Hartvigsen G, Ellingsen G, Bellika JG. Bridging the Gap between Patients' Expectations and General Practitioners' Knowledge through Disease Surveillance. Stud Health Technol Inform 2009;150:423-7.
60. Horst MA, Coco AS. Observing the spread of common illnesses through a community: using Geographic Information Systems (GIS) for surveillance. J Am Board Fam Med 2010;23:32-41.
61. Bellika JG, Sue H, Bird L, Goodchild A, Hasvold T, Hartvigsen G. Properties of a federated epidemiology query system. Int J Med Inform 2007;76:664-76.

Appendices/ Supplementary material

Appendix chapter 1

Syndrome definitions for the included medical registries.

Based on a literature search and a syndrome list as selected by the CDC (Centers for Disease Control and Prevention, USA) we selected syndromes that were expected to reflect the clinical presentations of both high-threat and common pathogens [1-3]. By this approach, we included respiratory, gastro-enteritis and neurological syndromes for evaluation. Table 1 shows high-threat pathogens and/or diseases that can present with any of these clinical syndromes.

Then, general respiratory, gastro-enteritis and neurological syndromes were defined for each medical registry in our project (except for the absenteeism registry which contained no medical information). In this chapter 1 appendix we only describe the general gastro-enteritis and neurological syndromes for the medical registries included in the project. The general respiratory (and additional) syndromes are described in detail in the appendices of chapters 2 to 5.

To define the syndromes, we used the ICD-9-CM *(International Classification of Diseases, 9th revision, Clinical Modification)* codes as selected by the CDC [3]. We selected both the codes for general symptoms and diagnoses ('category 1' in CDC-list) and the codes for specific pathogens diagnoses ('category 3' in CDC-list). For the hospital registry (see Table 2a and 2b) we used these syndrome codes with some minor adaptations for the Dutch version of ICD-9-CM. For the mortality registry (see Table 3a and 3b) the ICD-9-CM-codes were converted into ICD-10 *(International Classification of Diseases, 10th revision)* codes using the WHO ICD-9/ICD-10 translation list and expert opinion if necessary (ICD-9/ICD-10 Translator, see http://www.who.int/classifications/en/). For the GP registry (see Table 4a and 4b), ICPC *(International Classification of Primary Care)* codes were included in a gastro-enteritis/neurological syndrome by expert opinion, guided by the CDC syndrome case definitions.

For the pharmacy registry, we defined a gastro-enteritis syndrome by including all subcategories of the ATC *(Anatomical Therapeutic Chemical Classification System)* code A07 ("antidiarrheals, intestinal anti-inflammatory/anti-infective agents"), see Table 5. We did not investigate a neurological syndrome for the pharmacy data.

For the laboratory submissions registry we defined a gastro-enteritis syndrome by including all submissions for specific diagnostics on faeces material, but only if the requested diagnostics were likely to be related to gastro-enteral symptoms (according to expert opinion, see list of diagnostics in Table 6). For a neurological syndrome definition based on the laboratory submissions data we included all submissions for microbiological diagnostics on cerebrospinal fluids.

Reference List

1. Wagner MM, Dato V, Dowling JN, Allswede M. Representative threats for research in public health surveillance. J Biomed Inform 2003;36:177-88.
2. 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.
3. Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef/word/syndromedefinitions.doc, accessed at May 28th 2010.

Table 1. Examples of high threat pathogens that can cause a respiratory, gastro-enteritis or neurological syndrome [1-3].

	Respiratory syndrome	Gastro-enteritis syndrome	Neurological syndrome
Category A	*B. anthracis* (inhalational anthrax) *Y. pestis* (pneumonic plague) *F. tularensis* (tularemia)	*B. anthracis* (gastrointestinal anthrax)	
Category B	*C. burnetii* (Q fever) *C. psittaci* (Psittacosis)	Food safety threats (e.g., *Salmonella spp., Escherichia coli O157:H7*) Water safety threats (e.g., *Vibrio cholerae, Cryptosporidium parvum*)	Alphaviruses (VEE, EEE, WEE)
Category C	(Pandemic) influenza, SARS		West-Nile virus and other flaviviruses

Table 2. ICD-9-CM codes for the gastro-enteritis (a) and neurological syndrome (b) in hospital data.*

2a. Gastro-enteritis syndrome.

ICD-9-CM code	Description
001.0	Cholera due to *Vibrio cholerae*
001.1	Cholera due to *Vibrio cholerae el tor*
001.9	Cholera, unspecified
003.0	*Salmonella* gastroenteritis
003.20	Localized S*almonella* infection unspecified
003.29	Other localized *Salmonella* infections
003.8	Other specified *Salmonella* infections
003.9	*Salmonella* infection, unspecified
004.0	*Shigella dysenteriae*
004.1	*Shigella flexneri*
004.2	*Shigella boydii*
004.3	*Shigella sonnei*
004.8	Other specified *Shigella* infections
004.9	Shigellosis, unspecified
005.0	Staphylococcal food poisoning
005.2	Food poisoning due to *Clostridium perfringens* [C. welchii]
005.3	Food poisoning due to other Clostridia
005.4	Food poisoning due to *Vibrio parahaemolyticus*
005.8	Other bacterial food poisoning
005.9	Food poisoning, unspecified
006.0	Acute amebic dysentery without mention of abscess
006.8	Amebic infection of other sites
006.9	Amebiasis, unspecified
007.0	Balantidiasis
007.1	Giardiasis
007.2	Coccidiosis
007.3	Intestinal trichomoniasis
007.8	Other specified protozoal intestinal diseases
007.9	Unspecified protozoal intestinal disease
008.0	Intestinal infection due to *Escherichia coli* [*E. coli*]
008.1	Enteritis due to Arizona group of paracolon bacilli
008.2	Enteritis due to *Aerobacter aerogenes* enteritis
008.3	Enteritis due to *Proteus (mirabilis) (morganii)*

008.41	Intestinal infection due to *Staphylococcus*
008.49	Intestinal infection due to other gram-negative bacteria
008.5	Bacterial enteritis, unspecified
008.6	Enteritis due to specified virus
008.8	Enteritis due to other organism, not elsewhere classified
009.0	Infectious colitis, enteritis, and gastroenteritis
009.1	Colitis, enteritis, and gastroenteritis of presumed infectious origin
009.2	Infectious diarrhea
009.3	Diarrhea of presumed infectious origin
021.1	Enteric tularemia
022.2	Gastrointestinal anthrax
078.82	Epidemic vomiting syndrome
088.0	Intestinal infection due to e. coli unspecified
127.0	Ascariasis
127.1	Anisakiasis
127.2	Strongyloidiasis
127.3	Trichuriasis
127.4	Enterobiasis
127.5	Capillariasis
127.6	Trichostrongyliasis
127.7	Other specified intestinal helminthiasis
127.8	Mixed intestinal helminthiasis
127.9	Intestinal helminthiasis, unspecified
129	Intestinal parasitism, unspecified
535.0	Acute gastritis
535.4	Other specified gastritis
535.5	Unspecified gastritis and gastroduodenitis
535.6	Duodenitis
536.2	Persistent vomiting
555.0	Regional enteritis of small intestine
555.1	Regional enteritis of large intestine
555.2	Regional enteritis of small intestine with large intestine
558.2	Toxic gastroenteritis and colitis
558.9	Other and unspecified noninfectious gastroenteritis and colitis
567.0	Peritonitis in infectious diseases classified elsewhere
569.9	Unspecified disorder of intestine
787.0	Nausea and vomiting
787.3	Flatulence, eructation, and gas pain
787.4	Visible peristalsis
787.9	Other symptoms involving digestive system

2b. Neurological syndrome.

ICD-9-CM code	Description
003.21	*Salmonella* meningitis
036.0	Meningococcal meningitis
036.1	Meningococcal encephalitis
036.2	Meningococcemia
036.89	Infection, meningococcal nec
036.9	Meningococcal infection, unspecified
047.0	Meningitis due to coxsackie virus

047.1	Meningitis due to echo virus
047.8	Other specified viral meningitis
047.9	Unspecified viral meningitis
048	Other enterovirus diseases of central nervous system
049.0	Lymphocytic choriomeningitis
049.1	Meningitis due to adenovirus
049.8	Other specified non-arthropod-borne viral diseases of central nervous system
049.9	Unspecified non-arthropod-borne viral diseases of central nervous system
052.0	Postvaricella encephalitis
053.0	Herpes zoster with meningitis
053.10	Herpes zoster with unspecified nervous system complication
054.3	Herpetic meningoencephalitis
054.72	Hsv, meningitis
055.0	Postmeasles encephalitis
056.00	Rubella with unspecified neurological complication
056.01	Encephalomyelitis due to rubella
056.09	Rubella with other neurological complications
061	Dengue
062.0	Japanese encephalitis
062.1	Western equine encephalitis
062.2	Eastern equine encephalitis
062.3	St. louis encephalitis
062.4	Australian encephalitis
062.5	California virus encephalitis
062.8	Other specified mosquito-borne viral encephalitis
062.9	Mosquito-borne viral encephalitis, unspecified
063.0	Russian spring-summer [taiga] encephalitis
063.1	Louping ill
063.2	Central european encephalitis
063.8	Other specified tick-borne viral encephalitis
063.9	Tick-borne viral encephalitis, unspecified
064	Viral encephalitis transmitted by other and unspecified arthropods
066.4	West nile fever
071	Rabies
072.1	Mumps meningitis
072.2	Mumps encephalitis
084.9	Other pernicious complications of malaria
086.2	Chagas' disease without mention of organ involvement
086.3	Gambian trypanosomiasis
086.4	Rhodesian trypanosomiasis
086.5	African trypanosomiasis, unspecified
091.81	Acute syphil meningitis
098.82	Gonoccocal, meningitis
100.81	Leptospiral infections,meningitis (aseptic)
114.2	Coccidioidal meningitis
115.01	Histoplasmosis meningitis
115.11	*Histoplasma duboisii*, meningitis
115.91	Histoplasmosis, unspec, meningitis
117.5	Cryptococcosis

130.0	Meningoencephalitis due to toxoplasmosis
136.2	Specific infections by free-living amebae
320.0	Hemophilus meningitis
320.1	Pneumococcal meningitis
320.2	Streptococcal meningitis
320.3	Staphylococcal meningitis
320.7	Meningitis in other bacterial diseases classified elsewhere
320.8	Meningitis due to other specified bacteria
320.9	Meningitis due to unspecified bacterium
321.0	Cryptococcal meningitis
321.1	Meningitis in other fungal diseases
321.2	Meningitis due to viruses not elsewhere classified
321.3	Meningitis due to trypanosomiasis
321.4	Meningitis in sarcoidosis
321.8	Meningitis due to other nonbacterial organisms classified elsewhere
322.0	Nonpyogenic meningitis
322.1	Eosinophilic meningitis
322.9	Meningitis, unspecified
323.0	Encephalitis in viral diseases classified elsewhere
323.1	Encephalitis in rickettsial diseases classified elsewhere
323.2	Encephalitis in protozoal diseases classified elsewhere
323.4	Other encephalitis due to infection classified elsewhere
323.5	Encephalitis following immunization procedures
323.6	Postinfectious encephalitis
323.7	Toxic encephalitis
323.8	Other causes of encephalitis
323.9	Unspecified cause of encephalitis
348.3	Encephalopathy, unspecified
781.6	Meningismus

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

Table 3. ICD-10 codes for the gastro-enteritis (a) and neurological syndrome (b) in mortality data.*

3a. Gastro-enteritis syndrome.

ICD-10 code	Description
A000	Cholera due to *Vibrio cholerae* 01, biovar cholerae
A001	Cholera due to *Vibrio cholerae* 01, biovar eltor
A009	Cholera, unspecified
A020	*Salmonella enteritis*
A022	Localized salmonella infections
A028	Other specified salmonella infections
A029	*Salmonella* infection, unspecified
A030	Shigellosis due to *Shigella dysenteriae*
A031	Shigellosis due to *Shigella flexneri*
A032	Shigellosis due to *Shigella boydii*
A033	Shigellosis due to *Shigella sonnei*
A038	Other shigellosis
A039	Shigellosis, unspecified

A040	Enteropathogenic *Escherichia coli* infection
A041	Enterotoxigenic *Escherichia coli* infection
A042	Enteroinvasive *Escherichia coli* infection
A043	Enterohaemorrhagic *Escherichia coli* infection
A044	Other intestinal *Escherichia coli* infections
A045	*Campylobacter* enteritis
A046	Enteritis due to *Yersinia enterocolitica*
A047	Enterocolitis due to *Clostridium difficile*
A048	Other specified bacterial intestinal infections
A049	Bacterial intestinal infection, unspecified
A050	Foodborne staphylococcal intoxication
A052	Foodborne *Clostridium perfringens* [*Clostridium welchii*] intoxication
A053	Foodborne *Vibrio parahaemolyticus* intoxication
A054	Foodborne *Bacillus cereus* intoxication
A058	Other specified bacterial foodborne intoxications
A059	Bacterial foodborne intoxication, unspecified
A060	Acute amoebic dysentery
A061	Chronic intestinal amoebiasis
A063	Amoeboma of intestine
A069	Amoebiasis, unspecified
A070	Balantidiasis
A071	Giardiasis [lambliasis]
A072	Cryptosporidiosis
A073	Isosporiasis
A078	Other specified protozoal intestinal diseases
A079	Protozoal intestinal disease, unspecified
A080	Rotaviral enteritis
A081	Acute gastroenteropathy due to Norwalk agent
A082	Adenoviral enteritis
A083	Other viral enteritis
A084	Viral intestinal infection, unspecified
A085	Other specified intestinal infections
A09	Diarrhoea and gastroenteritis of presumed infectious origin
A213	Gastrointestinal tularaemia
A22	Anthrax
A881	Epidemic vertigo
B770	Ascariasis with intestinal complications
B779	Ascariasis, unspecified
B780	Intestinal strongyloidiasis
B787	Disseminated strongyloidiasis
B789	Strongyloidiasis, unspecified
B79	Trichuriasis
B80	Enterobiasis
B810	Anisakiasis
B811	Intestinal capillariasis
B812	Trichostrongyliasis
B813	Intestinal angiostrongyliasis
B814	Mixed intestinal helminthiases
B818	Other specified intestinal helminthiases

B820	Intestinal helminthiasis, unspecified
B829	Intestinal parasitism, unspecified
K29	Gastritis and duodenitis
K500	Crohn's disease of small intestine
K501	Crohn's disease of large intestine
K508	Other Crohn's disease
K521	Toxic gastroenteritis and colitis
K529	Noninfective gastroenteritis and colitis, unspecified
K639	Disease of intestine, unspecified
K670	Chlamydial peritonitis (A74.8)
K671	Gonococcal peritonitis (A54.8)
K672	Syphilitic peritonitis (A52.7)
K673	Tuberculous peritonitis (A18.3)
K678	Other disorders of peritoneum in infectious diseases classified elsewhere
K929	Disease of digestive system, unspecified
K930	Tuberculous disorders of intestines, peritoneum and mesenteric glands (A18.3)
R11	Nausea and vomiting
R14	Flatulence and related conditions
R19	Other symptoms and signs involving the digestive system and abdomen
T629	Noxious substance eaten as food, unspecified

3b. Neurological syndrome.

ICD-10 code	Description
A390	Meningococcal meningitis (G01*)
A392	Acute meningococcaemia
A394	Meningococcaemia, unspecified
A398	Other meningococcal infections
A399	Meningococcal infection, unspecified
A548	Other gonococcal infections
A800	Acute paralytic poliomyelitis, vaccine-associated
A801	Acute paralytic poliomyelitis, wild virus, imported
A802	Acute paralytic poliomyelitis, wild virus, indigenous
A803	Acute paralytic poliomyelitis, other and unspecified
A809	Acute poliomyelitis, unspecified
A818	Other slow virus infections of central nervous system
A820	Sylvatic rabies
A821	Urban rabies
A829	Rabies, unspecified
A830	Japanese encephalitis
A831	Western equine encephalitis
A832	Eastern equine encephalitis
A833	St Louis encephalitis
A834	Australian encephalitis
A835	California encephalitis
A836	Rocio virus disease
A838	Other mosquito-borne viral encephalitis
A839	Mosquito-borne viral encephalitis, unspecified
A840	Far Eastern tick-borne encephalitis [Russian spring-summer encephalitis]

A841	Central European tick-borne encephalitis
A848	Other tick-borne viral encephalitis
A849	Tick-borne viral encephalitis, unspecified
A850	Enteroviral encephalitis (G05.1*)
A851	Adenoviral encephalitis (G05.1*)
A852	Arthropod-borne viral encephalitis, unspecified
A858	Other specified viral encephalitis
A86	Unspecified viral encephalitis
A87	Viral meningitis
A888	Other specified viral infections of central nervous system
A89	Unspecified viral infection of central nervous system
A90	Dengue fever [classical dengue]
B003	Herpesviral meningitis (G02.0*)
B004	Herpesviral encephalitis (G05.1*)
B010	Varicella meningitis (G02.0*)
B011	Varicella encephalitis (G05.1*)
B020	Zoster encephalitis (G05.1*)
B021	Zoster meningitis (G02.0*)
B022	Zoster with other nervous system involvement
B050	Measles complicated by encephalitis (G05.1*)
B060	Rubella with neurological complications
B261	Mumps meningitis (G02.0*)
B262	Mumps encephalitis (G05.1*)
B341	Enterovirus infection, unspecified
B38	Coccidioidomycosis
B451	Cerebral cryptococcosis
B457	Disseminated cryptococcosis
B458	Other forms of cryptococcosis
B459	Cryptococcosis, unspecified
B500	*Plasmodium falciparum* malaria with cerebral complications
B509	*Plasmodium falciparum* malaria, unspecified
B520	*Plasmodium malariae* malaria with nephropathy
B54	Unspecified malaria
B569	African trypanosomiasis, unspecified
B57	Chagas' disease
B580	Toxoplasma oculopathy
B582	Toxoplasma meningoencephalitis (G05.2*)
B602	Naegleriasis
B832	Angiostrongyliasis due to Parastrongylus cantonensis
G00	Bacterial meningitis, not elsewhere classified
G01	Meningitis in bacterial diseases classified elsewhere
G02	Meningitis in other infectious and parasitic diseases classified elsewhere
G03	Meningitis due to other and unspecified causes
G040	Acute disseminated encephalitis
G042	Bacterial meningoencephalitis and meningomyelitis, not elsewhere classified
G048	Other encephalitis, myelitis and encephalomyelitis
G049	Encephalitis, myelitis and encephalomyelitis, unspecified
G05	Encephalitis, myelitis and encephalomyelitis in diseases classified elsewhere
G062	Extradural and subdural abscess, unspecified
G07	Intracranial and intraspinal abscess and granuloma in diseases classified elsewhere

G373	Acute transverse myelitis in demyelinating disease of central nervous system
G374	Subacute necrotizing myelitis
G92	Toxic encephalopathy
G96	Other disorders of central nervous system
R29	Other symptoms and signs involving the nervous and musculoskeletal systems

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

Table 4. ICPC codes for the gastro-enteritis (a) and neurological syndrome (b) in GP consultations data.*

4a. Gastro-enteritis syndrome.

ICPC code	Description
D01	Abdominal pain/cramps general
D02	Abdominal pain epigastric
D03	Heartburn
D04	Rectal/anal pain
D06	Abdominal pain localized other
D08	Flatulence/gas/belching
D09	Nausea
D10	Vomiting
D11	Diarrhoea
D14	Haematemesis/vomiting blood
D16	Rectal bleeding
D18	Change faeces/bowel movements
D24	Abdominal mass not otherwise specified
D25	Abdominal distension
D29	Digestive symptom/complaint other
D70	Gastrointestinal infection
D73	Gastroenteritis presumed infection
D87	Stomach function disorder
D99	Disease digestive system, other

4b. Neurological syndrome.

ICPC code	Description
A07	Coma
N01	Headache
N03	Pain face
N07	Convulsion/seizure
N19	Speech disorder
N29	Neurological symptom/complt. other
N70	Poliomyelitis
N71	Meningitis/encephalitis
N72	Tetanus
N73	Neurological infection other
N89	Migraine
N99	Neurological disease, other
P20	Memory disturbance

*ICPC, International Classification of Primary Care

Table 5. ATC codes for the gastro-enteritis syndrome in pharmacy data.*
(All 5-digit ATC codes listed in the table are subcategories of the 3-digit ATC code A07: antidiarrheals, intestinal anti-inflammatory/anti-infective agents.)

ATC-5 code	Description
A07AA	Antibiotics
A07AB	Sulfonamides
A07AC	Imidazole derivatives
A07AX	Other intestinal anti-infectives
A07BA	Charcoal preparations
A07BB	Bismuth preparations
A07BC	Other intestinal adsorbents
A07CA	Oral rehydration salt formulations
A07DA	Antipropulsives
A07EA	Corticosteroids acting locally
A07EB	Antiallergic agents, excluding corticosteroids
A07EC	Aminosalicylic acid and similar agents
A07FA	Antidiarrheal micro-organisms
A07XA	Other antidiarrheals

*ATC, Anatomical Therapeutic Chemical Classification System

Table 6. Laboratory test requests for stool samples that were included in the gastro-enteris syndrome for laboratory submissions.

Subjects of stool-sample test requests
Adeno 31 virus
Adeno 40 virus
Adeno 41 virus
Adeno virus
Aeromonas caviae
Aeromonas hydrophila
Aeromonas sobria
Aeromonas species
Amoeba
Antigen *Cryptosporidium* species
Antigen *Entamoeba histolytica*
Antigen *Giardia lamblia*
Arcobacter butzleri
Ascaris lumbricoides
Bacillus cereus
Bacillus species
Bacterium
Balantidium coli
Blastocystis hominis
Calici noro virus
Campylobacter species (and sub-species and –types)
Clonorchis sinensis
Clostridium clostridiiforme

Clostridium difficile
Clostridium perfringens
Clostridium species
Cryptococcus neoformans
Cryptococcus species
Cryptosporidium parvum
Cryptosporidium species
Cyclospora cayetenesis
Cyclospora species
Cysts *Blastocystis hominis*
Cysts *Cryptosporidium* species
Cysts *Entamoeba polecki*
Dientamoeba fragilis
DNA *Entamoeba dispar*
DNA *Entamoeba histolytica*
Eggs *Diphyllobotrium latum*
Eggs *Trichuris trichiura*
Endolimax nana
Entamoeba coli
Entamoeba dispar
Entamoeba hartmanni
Entamoeba histolytica
Entamoeba histolytica dispar
Entamoeba histolytica hematofaag
Entamoeba species
Enterobacter species
Enterococcus faecium
Enterococcus species Vancomycine resistent
Escherichia coli (and subtypes)
Escherichia species
Giardia
Giardia lamblia
Iodamoeba butschlii
Isospora belli
Listeria monocytogenes
Listeria species
Microsporidia
Eggs *Ancylostoma/Necator*
Mycobacterium avium complex
Norwalk-like virus / norovirus
Parasite
Plesiomonas shigelloides
Prototheca zopfii(alg)
Protozoa
Rhabditiforme larven *Strongyloides* species
Rotavirus
Salmonella species (and sub-species and -types)
Shigella species (and all subtypes)

Small round structured viruses
Strongyloides species
Strongyloides stercoralis
Chilomastix mesnili
Entamoeba hartmanni
Vibrio cholerae
Vibrio parahaemolyticus
Vibrio species
Virus
Worms
Yersinia bercovieri
Yersinia enterocolitica (and subtypes)
Yersinia frederiksenii
Yersinia kristensenii
Yersinia mollaretti
Yersinia rohdei
Yersinia species

Appendix chapter 2

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 (http://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 http://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 Page 2 of 11 (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 or 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 tularaemia
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 capsulati
B392	Pulmonary histoplasmosis capsulati, unspecified
B393	Disseminated histoplasmosis capsulati
B394	Histoplasmosis capsulati, 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 apnoea
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	Haemothorax
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	Haemoptysis
R049	Haemorrhage 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/dyspnoea
R03	Wheezing
R04	Breathing problem, other
R05	Cough
R07	Sneezing/nasal congestion
R21	Throat symptom/complaint
R24	Haemoptysis
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 codes (5 digits) 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:

$S_t = b0 + b_1 P_{A,T+x} + b2 P_{B,t+y} + \ldots + R_t$

S = level of a respiratory syndrome

t = time in weeks

PA/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 three 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.

Appendix chapter 3

Appendix A: Detailed syndrome definitions for respiratory hospitalizations and mortality.

For the hospitalizations, we used discharge and secondary diagnoses on date of hospitalization from the Dutch National Medical Register (LMR, 99% coverage over 1999-2004, coded in Dutch version of ICD-9-CM). For the mortality we used primary cause of death, as well as complicating and other additional causes of death from Statistics Netherlands (CBS, http://www.cbs.nl, 1999-2004, 100% coverage, by date of death, coded in ICD-10 (Internat. Classification of Diseases, 10th revision))

To define respiratory syndromes, we used the ICD-9-CM *(International Classification of Diseases, 9th revision, Clinical Modification)* codes as selected by the CDC (Centers for Disease Control and Prevention, USA, www.bt.cdc.gov/surveillance/syndromedef).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 (Table A1) we used these syndrome codes with some minor adaptations for the Dutch version of ICD-9-CM. For the mortality data (table A2) the ICD-9-CM-codes were converted into ICD-10 codes using the WHO ICD-9/ICD-10 translation list and expert opinion if necessary (ICD-9/ICD-10 Translator, see http://www.who.int/classifications/en/).

Table A1. ICD9-CM codes for the respiratory syndrome in hospital data.

ICD-9-CM	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
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

Table A2. ICD-10 codes for the respiratory syndrome in mortality data.

ICD-10	Description
A202	Pneumonic plague
A212	Pulmonary tularaemia
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 capsulati
B392	Pulmonary histoplasmosis capsulati, unspecified
B393	Disseminated histoplasmosis capsulati
B394	Histoplasmosis capsulati, 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 apnoea
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	Haemothorax
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	Haemoptysis
R049	Haemorrhage 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

Appendix B: details on the regression models

Regression model variables

We constructed a GEE regression model with a poisson error and an identity link. For this, an additive model for the incidence was multiplied by the population size to obtain a model for the counts. Thus we could model proportional associations between the explanatory variables and the poisson distributed outcome variables, taking into account changes in population size.

Model for the counts:

$$C(t) = Pop(t) * I(t) + R(t)$$

Model for the incidence:

$$I(t) = b_0 + b_1 t + b_2 ILI(t+n_2) + b_3 P_3(t+n_3) + b_4 P_4(t+n_4) + \ldots + b_z P_z(t+n_z)$$

$C(t)$	the counts of a respiratory syndrome at time t (hospitalizations or mortality)
$Pop(t)$	the population size at time t
$I(t)$	the incidence of a respiratory syndrome at time t (hospitalizations or mortality)
$R(t)$	the residual of the count model
t	time in weeks over the whole study period (1-313 weeks)
b_0	regression coefficient describing constant basic syndrome level (lowered in summertime by using a dummy (0/1) variable to indicate the summer)
b_1-b_z	regression coefficients
$ILI(t)$	ILI-incidence at time t
$P_i(t)$	incidence of i-th respiratory pathogen at time t detected in the laboratory, i=2, 3, 4, …,z
n_i	lag time in weeks for ILI or the i-th pathogen, $-5 \leq n_i \leq 5$ weeks.

(Note that for interpretation purposes, we transformed the count model outcomes into incidences, see Figure 2.)

First a forward stepwise regression approach was used to construct a generalized linear model, each step selecting the lagged (-/+5 weeks) 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 ($P \leq 0.05$). Negative associations were excluded to avoid biological implausible associations in the models between the pathogens and the syndromes (e.g. negative associations between enteroviruses which peak in summer vs respiratory syndromes which peak in winter). We then – guided by periodograms of the residuals – added seasonal terms to the models (sine(k2πweek/52) and cosine(k2πweek/52), k=1,2,3,4) to correct for seasonal variation and used GEEs[16] to include autocorrelation in the residuals in the models. For the GEEs we defined each influenza year as a subject (July 1st-June 30th). Since we used GEE models with autocorrelated residuals, confidence intervals of prediction were not available, therefore we defined the one-sided 95% upper limit of the models as the 95 percentile of the residual divided by the square root of the predicted value. By plotting this upper limit together with the actual observations (Figure 2), the 5% most extreme upper observations can be recognized.

We then replaced the constant regression coefficient for ILI by different regression coefficients for separate years (between July 1st-June 30st) (again multiplying the model for the incidence by the populations size, see formulas above):

$$I(t) = b_0 + \sum_{k=1}^{6} b_{1k}\, ILI_k(t+n_1) + b_2 P_2(t+n_2) + b_3 P_3(t+n_3) + \ldots b_z P_z(t+n_z)$$

The regression coefficients and variables for ILI in this model are:

b_{1k}	regression coefficient for ILI-incidence in year k, k+1,…,6
k	first year of the analysis period (1999)
$ILI_k(t)$	ILI-incidence in year k ($ILI_k(t) = 0$ for t outside year k).

The annual regression coefficients for ILI (b_{1k}) in this model describe the association between hospitalizations or mortality and incidence of ILI by year.

In Table B1 the model outcomes for the models with a constant ILI regression coefficient are presented. The lag time (in weeks) is indicated, that showed optimal fit between hospitalizations/mortality time series and lagged ILI/pathogen counts included in the regression model. E.g.: according to the model, the trend in ILI for the 5-19 years of age precedes the hospitalizations in the 5-19 years of age with 1 week. For the time series on pathogens no information on age was available. This possibly explains the negative lags that were included in the models, since these time series reflect increased pathogen activity over all age categories, while the peak activity may differ in time between age categories. Some of the included pathogen variables were not significant (NS) anymore after correcting for seasonal variation and autocorrelation. ("-" = pathogen not included in the model; n/a indicates that the analyses were not performed for mortality in the 0-19 years of age due to sporadic counts in that category, see Figure 3a-b for the estimated annual ILI regression coefficients.)

Table B1. Estimated model regression coefficients: respiratory hospitalizations and mortality explained by ILI and respiratory pathogens with adjustment for seasonal trends (sine/cosine terms) and autocorrelation, 1999-2005.

		0-4 years		5-19 years		20-64 years		>=65 years	
Hospita-lizations		**Estimates**	**lag**	**Par.estimate**	**lag**	**Par.estimate**	**Lag**	**Par.estimate**	**lag**
	ILI	1.48	-1	0.40	1	0.62	0	3.52	0
	RSV	1.75	1	-	-	-	-	NS	-2
	Rhinovirus	-	-	NS	2	0.15	-4	0.63	-5
	Adenovirus	0.93	-1	NS	0	NS	0	0.63	-4
	Parainfluenza	NS	3	0.09	4	-	-	-	-
	B. pertussis	0.04	4	-	-	-	-	-	-
	M. Pneumoniae	-	-	-	-	-	-	-	-
Mortality									
	ILI	n/a	n/a	n/a	n/a	0.07	-1	2.90	-1
	RSV	n/a	n/a	n/a	n/a	NS	-5	NS	-4
	Rhinovirus	n/a	n/a	n/a	n/a	-	-	-	-
	Adenovirus	n/a	n/a	n/a	n/a	0.02	-4	0.41	-4
	Parainfluenza	n/a	n/a	n/a	n/a	-	-	-	-
	B. pertussis	n/a	n/a	n/a	n/a	-	-	-	-
	M. Pneumoniae	n/a	n/a	n/a	n/a	-	-	0.29	-5

Appendix chapter 4

Appendix S1: Detailed syndrome definition for hospitalizations with lower-respiratory infection syndrome.

We used discharge and secondary diagnoses on date of hospitalization from the Dutch National Medical Register (LMR, 99% coverage over 1999-2004, 80 % coverage over 2005-2006 16 million pop., coded in Dutch version of ICD-9-CM).

To define a lower-respiratory infection syndrome, we selected ICD-9-CM (International Classification of Diseases, 9th revision, Clinical Modification) codes for any kind of lower-respiratory infection from the respiratory syndrome codes-list as selected by the CDC (Centers for Disease Control and Prevention, USA, http://www.bt.cdc.gov/surveillance/syndromedef/). See Table 1 below for all selected lower-respiratory infection codes. We selected these codes as a subset from the set of 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). Finally these syndrome codes were slighty adapted for the Dutch version of ICD-9-CM.

Table 1. ICD-9-CM codes for lower-respiratory infection syndrome in hospitalization data.

ICD-9-CM	Description
003.22	*Salmonella* pneumonia
020.3	Primary pneumonic plague
020.4	Secondary pneumon plague
020.5	Pneumonic plague not otherwise specified
021.2	Pulmonary tularemia
022.1	Pulmonary anthrax
031.0	Mycobacteria, pulmonary
052.1	Varicella with pneumonia
055.1	Postmeasles pneumonia
073.0	Ornithosis, with pneumonia
114.0	Primary coccidioidomycosis (lung)
114.9	Coccidioidomycosis not otherwise specified
115.05	*Histoplasma capsulatum* pneumonia
115.15	*Histoplasma duboisii* pneumonia
115.95	Histoplasmosis pneumonia
116.0	Blastomycosis
116.1	Paracoccidioidomycosis
117.3	Pulmonary aspergillosis
130.4	Toxoplasma pneumonitis
136.3	Pneumocystosis
466.0	Bronchitis acute
466.1	Acute bronchiolitis
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 bact. not elsewhere classified
482.9	Pneumonia, bacterial 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 infectious disease not elsewhere classified
485	Bronchopneumonia organism unspec
486	Pneumonia, organism not otherwise specified
487.0	Influenza with pneumonia
490	Bronchitis not otherwise specified
511.1	Pleurisy with effusion, with mention of a bacterial cause other than tuberculosis

Appendix S2: Details on space-time analyses and Satscan settings

For detection of space-time clusters in the LRI-hospitalizations data, we used a space-time permutation scan statistic that compares observed and expected numbers of cases in circular areas with variable radii in flexible time periods. A likelihood ratio is calculated for each space-time window, to indicate to what extent the rate of cases inside the area is higher than expected. Monte Carlo hypothesis testing is then used to indicate the significance level of specific space-time windows. As expected numbers are calculated from the geographic distribution of cases in the whole dataset this method does not require additional population-at-risk data, and population density and seasonal variation in the case data is automatically adjusted for [15].

We simulated a prospective surveillance by running the scan-statistic on data from the year preceding each time unit (day or week) in the analysis period. This way, weekly or daily space-time signals were generated. For analyses in 1999, we used data from 2000 as historical data (by generating a stand-in-dataset for 1998 based on data from 2000). For all analyses, we chose to use time aggregation windows of 7-days length, even for the daily analysis for 1999 and 2006. For these daily analyses, the 7-day windows shifted one day forward for each daily run. Thus we both limited the computation time and adjusted for day-of-week effects (both purely temporal and spatial day-of-week effects). To further limit the computation time - for the initial analyses - we chose a maximum spatial cluster-signal size of 40% population at risk (instead of default 50%, the cases are here the population) and a maximum temporal cluster-signal size of 21 weeks (to not only detect outbreaks that evolve in e.g. 1 or 2 weeks, but also more gradually evolving outbreaks).

As a covariate we included the age group (0-4, 5-19, 20-49, 50-64, >=65 years).

To measure the significance of the detected cluster-signals we used recurrence-intervals. The recurrence interval reflects how often a signal of the observed significance level would be observed by chance, assuming that analyses are repeated on a regular basis (e.g. daily/weekly) [18]. E.g. a signal with a recurrence interval

of 1 year would, on average, be observed every year. The recurrence interval can be extracted directly from the p-value: e.g. for daily analysis the recurrence interval in days can be calculated as "1/p-value" whereas for weekly analysis the recurrence interval in days is "7/p-value". This implies that a recurrence interval of 1 year corresponds with a p-value of 0.00274 for daily analyses and 0.0192 for weekly analyses.

A sensitivity analysis was used to evaluate the impact of time and spatial window settings on the number of space-time signals detected. For the initial analyses, as described above we put only minor constraints on the maximum temporal and spatial windows of the scan-statistic. We then repeated these weekly analyses with a temporal window of maximum 7 weeks and also with a spatial window of maximum 25 km radius (arbitrarily chosen). We then used the initial analyses results as reference for the analyses with restricted settings, and evaluated what signals and clusters from the initial analyses were still detected with the restricted settings (by assessing geographical overlap between signals generated at the same date). Restrictions on space and time will lead to (little) extra signals as well, as borderline significant signals as detected with restricted settings, might have been not significant with non restrictive settings - due to more adjustments for taking into account multiple testing - stemming from the many (more) potential cluster locations and sizes evaluated [15].

We also performed space-time scans on the ILI-data, to assess whether regional ILI-clusters might explain detected LRI-clusters. For the ILI-data prospective space-time scans were performed, with a Poisson-distributed number of events in a geographical area, according to a known underlying population-at-risk [30]. Here we chose to use population-at-risk data, because of variation in the number of GP-practices supplying data (e.g. during vacations). We adjusted for purely temporal clusters using the time-stratified randomization option of the Satscan software, to prevent concurring ILI-clusters in all regions, e.g. by seasonal variation. In contrast with the LRI-analyses we only used recurrence intervals of >= 1 year, we divided the Netherlands in 4 major regions as spatial input, we used an analysis period of 180 days back from each date the prospective scan was performed for, we took a maximum spatial cluster-signal size of 50% population at risk (default) and we used a maximum temporal cluster-signal size of 7 days (to only detect weekly cluster-signals corresponding with the weekly regional ILI-incidence fluctuations as reported in the Dutch Influenza News letter, http://www.virology.nl/files/new.pdf). No covariate (age) was used and spatially the clusters were limited to contain only 1 out of the 4 major regions.

Appendix chapter 5

Appendix A: Detailed case definition for hospitalizations with LRI, hepatitis, and/or endocarditis syndromes

We used discharge diagnoses and secondary diagnoses by date of hospitalization, provided by the Dutch National Medical Register (LMR). Coded in the Dutch version of ICD-9-CM, the Register had 80% coverage over 16 million population 2005-2007. We included all records on hospitalizations with discharge or secondary diagnosis indicative for clinical syndromes that might be attributed to Q fever, i.e. (1) LRI, (2) hepatitis, and/or (3) endocarditis. We also included ICD-9-codes for specific diseases other than Q fever, making it possible to distinguish alternative possible causes for detected space-time clusters (by the proportion of cluster patients with those specific discharge diagnoses).

Although a specific ICD-9 code for Q fever exists (0830), we found for our study period only 18 hospitalization records with Q fever as a primary and/or secondary diagnosis. Of these, 13 occurred in 2007, reflecting only a fraction of the known outbreak cases; 9 of these 18 hospitalizations were included in the LRI syndrome based on other registered diagnostic codes (e.g. pneumonia) and likewise 1 hospitalization was included in the hepatitis case definition. The remaining 8 hospitalizations were excluded, since they did not have diagnoses that indicated LRI, hepatitis, and/or endocarditis.

For the LRI syndrome, we used the same codes as selected and described in an earlier study [1]. Below, the selected ICD-9 codes are presented for the hepatitis and endocarditis syndromes.

Hepatitis

In Table A1, the selected ICD-9 codes for acute and/or chronic hepatitis are presented.

Table A1. ICD-9-CM codes for hepatitis in hospitalization data.

ICD-9-CM	Description
070	Viral hepatitis
070.0	Viral hepatitis A with hepatic coma
070.1	Viral hepatitis A without mention of hepatic coma
070.2	Viral hepatitis B with hepatic coma
070.3	Viral hepatitis B without mention of hepatic coma
070.4	Other specified viral hepatitis with hepatic coma
070.5	Other specified viral hepatitis without mention of hepatic coma
070.6	Unspecified viral hepatitis with hepatic coma
070.9	Unspecified viral hepatitis without mention of hepatic coma
072.71	Mumps hepatitis
091.62	Secondary syphilitic hepatitis
130.5	Hepatitis due to toxoplasmosis
571.4	Chronic hepatitis
573.1	Hepatitis in viral diseases classified elsewhere
573.2	Hepatitis in other infectious diseases classified elsewhere
573.3	Hepatitis, unspecified

Endocarditis

In Table A2, the selected ICD-9 codes for acute and/or chronic endocarditis are presented.

Table A2. ICD-9-CM codes for endocarditis in hospitalization data.

ICD-9-CM	Description
036.42	Meningococcal endocarditis
074.22	Coxsackie endocarditis
093.2	Syphilitic endocarditis
098.84	Gonococcal endocarditis
112.81	Candidal endocarditis
115.04	*Histoplasma capsulatum* endocarditis
115.14	*Histoplasma duboissii* endocarditis
115.94	Histoplasmosis endocarditis
391.1	Acute rheumatic endocarditis
421	Acute and subacute endocarditis
421.0	Acute and subacute bacterial endocarditis
421.1	Acute and subacute infective endocarditis in diseases classified elsewhere
421.9	Acute endocarditis, unspecified
424.9	Endocarditis, valve unspecified

Reference

1. Van den Wijngaard CC, et al. Syndromic surveillance for local outbreaks of lower-respiratory infections: would it work? PLoS ONE 2010;5:e10406.

Appendix B

Table B1 and B2 present detailed information of all detected LRI and hepatitis clusters. Details presented on each cluster are: the weekdate of the first cluster signal (column 2), the total number of cluster signals (with recurrence intervals ≥ 1 years or 5 years, column 3), the total episode of the cluster (from the earliest weekdate of all cluster-signal episodes to the latest weekdate, column 4), the lowest and the highest difference for a cluster signal between the number of observed and the number of expected cases (column 5), the lowest and the highest cluster-signal radius (column 6), all relevant cluster characteristics and circumstantial evidence (column 7), and whether or not a cluster that was possibly due to Q fever regarding the criteria in Figure 1 overlapped with positive-tested small-ruminant farms (column 8). Finally, in the last column it is concluded, based upon the information in column 7 and 8 and the criteria in Figure 1, whether Q fever seems a plausible cause for the clusters or that other causes seem more likely.

Table B1. Detailed description of all space-time clusters in hospitalizations with LRI in 2005-2007.

Cluster	1st signal	Total no. signals – Sign. level		Total episode	No. observed cases- no. expected		Radii in km	Cluster characteristics and /or circumstantial evidence	Spatial overlap* with positive- tested goat/ sheep farms†	Might the cluster have been caused by Q fever?
		1 yr	5 yr		Lowest	Highest				
1	30 Oct. 2004	3	8	8 Aug. 2004- 8 Jan. 2005	1994-1782 = 212	6513-6045 = 468	33-50	• All cluster signals had higher proportions ($P \leq 0.01$) of bronchitis/ bronchiolitis cases and patients 0-4 yr of age. • 1 cluster signal had higher proportions of unexplained pneumonia. • 7 cluster signals overlapped with regional ILI elevations.	n/a	Q fever seems **unlikely**. Likely other causes: RSV/influenza.
2‡	1 Jan. 2005	1	16	26 Dec. 2004-23 Apr. 2005	156-104 = 52	1012-827 = 185	33-40	• 16 cluster signals had higher proportions ($P \leq 0.01$) of bronchitis/ bronchiolitis cases and patients 0-4 yr of age. • 13 cluster signals overlapped with regional ILI elevations.	n/a	Q fever seems **unlikely**. Likely other causes: RSV/influenza.
3‡	30 Apr. 2005	1	8	26 Dec. 2004-25 June 2005	221-151 = 70	620-478 = 142	10-14	• 2 cluster signals had higher proportions ($P \leq 0.05$) of patients 20-49 yr of age. • All cluster signals overlapped with regional ILI elevations, but the cluster contains a relatively small area, whereas the regional ILI elevations are measured by dividing the Netherlands in 4 widespread areas.	4 goat farms: 1 tested positive in 2006, 1 in 2007, and 2 in 2008	Q fever seems a **plausible** cause.
4§	19 Nov. 2005	0	1	23 Oct. 2005-19 Nov. 2005	n/a	149-93 = 56	17	• The cluster signals overlapped with regional ILI elevations, but the cluster contains a relatively small area, whereas the regional ILI elevations are measured by dividing the Netherlands in 4 widespread areas (in addition, the national ILI incidence was below 3/10000 pop).	2 goat farms and 1 sheep farm: 1 goat farm tested positive in 2001, the other in 2007, and the sheep farm in 2008.	Q fever seems a **plausible** cause.
5§	26 Nov. 2005	0	5	24 July 2005-24 Dec. 2005	2766-2502 = 264	3708-3388 =320	51-53	• 2 cluster signals had higher proportions ($P \leq 0.05$) of hospitalizations with legionellosis as discharge diagnosis. • 5 cluster signals had higher proportions ($P \leq 0.01$) of patients 0-4 yr of age. • All cluster signals overlapped with regional ILI elevations, but the national ILI incidence was below 3/10000 pop.	n/a	Q fever seems **unlikely.** Likely other causes: RSV/influenza, possibly *Legionella* as well.
7	14 Jan. 2006	4	14	25 Dec. 2005-13 May 2006	220-153 = 67	2361-2130 = 231	28-59	• 8 cluster signals had higher proportions ($P \leq 0.05$) of bronchitis/ bronchiolitis cases (5 of these also at $P \leq 0.01$). • 16 cluster signals had higher proportions ($P \leq 0.01$) of patients 0-4 yr of age. • 1 cluster signal had higher proportions ($P \leq 0.05$) of patients 0-4 and 5-19 yr of age. • All cluster signals overlapped with regional ILI elevations. • A space-time cluster of 4 LD reports overlapped with the cluster; however, due to the large space-time window of the LRI cluster, this is possibly a coincidence (the LD cluster had a window of only 3 weeks).	n/a	Q fever seems **unlikely**. Likely other causes: RSV/influenza.
8	18 Mar. 2006	2	5	19 Feb. 2006- 6 May 2006	31-11= 20	1771-1574= 197	4-28	• 4 cluster signals had higher proportions of bronchitis/ bronchiolitis cases ($P \leq 0.01$), and 5 cluster signals had higher proportions of patients 0-4 yr of age ($p \leq 0.05$; 3 of these cluster signals also at $P \leq 0.01$).	n/a	Q fever seems **unlikely**. Likely other causes: RSV/influenza.
9	20 May 2006	1	0	14 May 2006-20 May 2006	n/a	5-0 = 5	3	• No special cluster signal characteristics observed.	2 goat farms (just next to the cluster area at the west side): both firstly tested positive in 2007.	Q fever seems a **plausible** cause.
10	15 July 2006	0	4	2 July 2006- 5 Aug. 2006	19-4 = 15	98-53 = 45	6-19	• All cluster signals had higher proportions ($P \leq 0.01$) of hospitalizations with LD as discharge diagnosis (ICD-9 4828). • All LRI cluster signals strongly overlapped both in space and time with a cluster of LD reports (max. 50 cases). • All cluster signals occurred in the already known episode and area of an LD outbreak in Amsterdam in 2006.	n/a	Q fever seems **unlikely**. Certainly another cause: *Legionella*.
11	16 Sept. 2006	6	0	16 July 2006-4 Nov. 2006	102-60 = 42	160-105 = 55	6	• All cluster signals had higher proportions ($P \leq 0.01$) of bronchitis/ bronchiolitis cases and patients 0-4 yr of age. • All cluster signals overlapped with regional ILI elevations, but the national ILI incidence was below 3/10000 pop.	n/a	Q fever seems **unlikely**. Likely another cause: RSV.

Cluster	1st signal	Total no. signals — Sign. level		Total episode	No. observed cases- no. expected		Radii in km	Cluster characteristics and /or circumstantial evidence	Spatial overlap* with positive- tested goat/ sheep farms†	Might the cluster have been caused by Q fever?
		1 yr	5 yr		Lowest	Highest				
13	13 Jan. 2007	5	15	17 Dec. 2006-26 May 2007	28-8 = 20	127-78 = 49	3-5	• All cluster signals had higher proportions ($P \leq 0.01$) of unexplained pneumonia cases. • 3 cluster signals had higher proportions of patients ≥65 yr of age, and 1 signal had patients 5-19 yr of age ($P \leq 0.05$). • 1 hospitalization was diagnosed with psittacosis pneumonia and 2 with aspergillosis (which was significantly different from the rest of the Netherlands, at $P \leq 0.01$ and 0.05, in that order). • 5 cluster signals overlapped with regional ILI elevations, but the national ILI incidence was below 3/10000 pop.	No overlap with positive-tested farms	Although this cluster did not exactly match the criteria for Q fever as a **possible** cause, we decided to assess whether it overlapped with reported Q-fever-abortion waves because of the consistently high proportion of unexplained pneumonia cases.
14	3 Feb. 2007	5	1	14 Jan. 2007-17 Mar. 2007	71-36 = 35	653-530 = 123	20-60	• All cluster signals had higher proportions ($P \leq 0.01$) of bronchitis/ bronchiolitis cases and patients 0-4 yr of age.	n/a	Q fever seems **unlikely**. Likely another cause: RSV/ influenza.
15	3 Mar. 2007	1	1	18 Feb. 2007-17 Mar. 2007	182-120 = 62	550-440 = 110	19-29	• 1 cluster signal had higher proportions ($P \leq 0.05$) of unexplained pneumonia cases. • Both cluster signals overlapped with regional ILI elevations.	9 goat farms: 2 goat farms tested positive in 2005, 3 goat farms in 2006, 2 in 2007, and 2 in 2008	Q fever seems a **plausible** cause (but regional influenza activity as well).
17	26 May 2007	1	0	13 May 2007-26 May 2007	n/a	11-1 = 10	6	• The cluster signal had higher proportion of patients 50-64 yr of age ($P \leq 0.05$). • The cluster signal overlapped with regional ILI elevations, but the national ILI incidence was below 3/10000 pop.	No overlap with positive-tested farms	Q fever seems **possible.**
18	26 May 2007	2	0	13 May 2007-2 June 2007	18-4 = 14	22-6 = 16	3	• The cluster signals had higher proportions of bronchitis/ bronchiolitis cases ($P \leq 0.01$). • The cluster signals had higher proportions of patients 50-64 yr of age ($P \leq 0.01$ and $P \leq 0.05$, in that order). • Abortion waves were reported at 1 sheep farm within close range of the cluster area in 2005-2008.	1 sheep farm (approximately 14 km south-west of the cluster center) firstly tested positive in 2006	Q fever seems a **plausible** cause.
20	15 Sept. 2007	2	0	19 Aug. 2007-19 Sept. 2007	44-19 = 25	62-31 = 31	10	• Both cluster signals had higher proportions ($P \leq 0.05$) of unexplained pneumonia cases. • The cluster signals overlapped in space and time with those of 6 Q fever patients between 2 Sept. and 29 Dec. 2007, but no Q-fever cases were reported for postal code areas that actually overlapped with the LRI cluster. • Abortion waves were reported at 1 goat farm within or at close range of the cluster area in 2005-2008.	1 goat farm that first tested positive in 2006	Q fever seems a **plausible** cause.

ILI, Influenza-like illness; RSV, respiratory syncytial virus; LD, Legionnaires' Disease.

* Spatial overlap with positive farms was assessed only if the cluster met the criteria for Q fever to be a possible cause.

† These were 26 Q fever positive-tested goat/sheep farms that voluntarily submitted abortion material in 2005-2008. Only 3 farms submitted the material after reporting of abortion waves became mandatory in 2008; one farm was retrospectively tested positive on material from 2001 (see Methods section).

‡ Clusters 2 and 3 overlapped in space and time but were not merged because a shared cause seemed unlikely, since cluster 3 contained a smaller geographical area and did not contain significantly higher proportions of patients 0-4 yr of age and bronchitis/bronchiolitis cases, whereas cluster 2 did.

§ Clusters 4 and 5 overlapped in space and time but were not merged because a shared cause seemed unlikely, since cluster 4 contained a much smaller space–time window and geographically overlapped with cluster 5 in only a borderline way.

Table B2. Detailed description of space-time clusters in hospitalizations with hepatitis in 2005-2007.

Set	Centre province	1st signal	Total no. signals – Sign. level		Total episode	No. observed – no. expected		Radii in km	Cluster characteristics and /or circumstantial evidence	Spatial overlap* with positive tested goat/ sheep farms†	Might the cluster have been caused by Q fever?
			1yr	5yr		lowest	highest				
1	UT	10 Feb. 2007	1	0	7 Jan. 2007-10 Feb. 2007	n/a	28-11=17	19	• The cluster signal overlaps in space and time with a cluster of 4 hepatitis C reports.	n/a	Q fever seems unlikely; possible cause is hepatitis C.
2	LI	12 May 2007	1	0	29 Apr. 2007-12 May 2007	n/a	10-2=8	26	-	2 goat farms, both first tested positive in 2007	Q fever seems a plausible cause.

* Spatial overlap with positive-tested farms was assessed only if the cluster met the criteria for Q fever to be a possible cause.

† These were 26 Q fever-positive-tested goat/sheep farms that voluntarily submitted abortion material in 2005-2008; only 3 farms submitted the material after reporting of abortion waves became mandatory in 2008; one farm retrospectively tested positive on material from 2001 (see Methods section).

Summary

Timely detection of emerging infectious diseases has become a priority in Public Health since the SARS epidemic in 2003, the bioterrorism attacks in 2001 and the ongoing threat of new infectious disease outbreaks. Early detection can be difficult, since patients infected with a high-threat pathogen like SARS or a new pandemic flu virus can come into the physician's office with symptoms typical of "common flu": fever, fatigue, some nausea, some coughing. Since on a daily basis many patients will present with such symptoms due to common diseases, and individual clinicians may only see one or a few extra patient(s) in case of a newly emerging epidemic, one cannot expect them to immediately recognize unusual disease. Furthermore, since for such common clinical symptoms often no laboratory diagnosis is made, traditional outbreak detection (based on astute clinicians and laboratory detection) could miss outbreaks by uncommon or unknown pathogens. Syndromic surveillance might reveal such blind spots of traditional surveillance by monitoring trends in non-specific clinical syndromes like "pneumonia by unknown cause", and by using information from very different sources like the number of cough medicine prescriptions. Therefore, prospective syndromic surveillance systems have been put in place in several countries, searching for unexpected elevations in various kinds of pre- or non-diagnostic data like absenteeism, medical telephone help-line calls, hospital emergency-department data or even internet searches and media reports. There has, however, been an ongoing debate about their effectiveness with especially concerns about the false alarms that these systems could generate. In this thesis, we studied the potential added value of syndromic surveillance, both for early warning detection of outbreaks and for tracking disease dynamics of (common) pathogens in the general population, including shifts in virulence or disease burden. For this, we included retrospective syndromic data from existing Dutch medical registries with data on work absenteeism, general-practitioner consultations, pharmacy dispensations, hospitalizations, laboratory submissions and mortality. Respiratory, gastro-enteritis and neurological syndromes were defined for these registries, since high-threat as well as common pathogens can cause these syndromes (chapter 1). We used time-series regression models that explained the syndrome time series by pathogen surveillance data, to estimate the contributions of individual pathogens on our syndromes. To evaluate whether local outbreaks can be detected in syndromic data, we applied space-time scan statistics to detect known historical outbreaks.

Syndromic data from medical registries reflects pathogen activity

Common pathogen activity was clearly reflected by the included syndromic data on respiratory, gastro-enteritis and neurological syndromes, with seasonal syndrome elevations concurring with elevations of specific pathogen counts (chapter 2 and 6). This implies that emerging pathogen activity could potentially be reflected in syndromic data as well. Respiratory syndromes best reflected known pathogen activity; in the regression models 68-86% of the syndrome variation was explained by the variation in pathogen counts, with influenza and RSV explaining most syndrome variation. For gastro-enteritis syndromes 29-85% of the variation was explained by known pathogen activity (rotavirus, *Shigella, Campylobacter* and *Salmonella*), and for a viral neurological syndrome 62% (enterovirus). Some syndromes were up to 5 weeks ahead of laboratory surveillance and some reflected pathogen activity that was not or incompletely detected by laboratory surveillance.

Disease burden and/or shifts in virulence can be monitored with syndromic surveillance

For influenza, using similar time-series regression models, we detected previously unknown annual shifts in the number of hospitalizations and deaths related to the number of influenza-like illness (ILI) cases, coinciding with shifts in antigenicity of circulating viruses (chapter 3). Thus combining analysis of syndromic and pathogen or disease specific data allows for better quantification of the impact of common pathogens on public health. Syndromic data was also used to detect a previously unknown burden of norovirus infections (chapter 6); a clear association with mild to severe morbidity and even deaths in the elderly was found. The latter was suspected but could not be assessed previously by any other routine surveillance.

Emerging disease outbreaks can be timely detected with syndromic surveillance

Surveillance of local syndrome elevations — if they are (still) too small to detect on the national level — might accelerate the detection of emerging outbreaks. To test this on known "gold standard" outbreaks, we used space-time scan statistics to detect known outbreaks of severe respiratory infections in syndromic hospital data (Legionnaires' disease in 1999 and 2006, chapter 4). Since only a modest overall number of alarms was generated in time (on average 5 detected clusters per year), this indicates that syndromic surveillance can also detect local outbreaks caused by uncommon or unknown pathogens. By the same approach, we also detected previously unknown disease clusters plausibly due to Q fever (chapter 5), illustrating that in some occasions syndromic surveillance can detect outbreaks that otherwise remain undetected. Real-time detection and investigation of these previously unknown clusters, may have led to the detection of increased Q-fever activity in the Netherlands well before its actual detection in 2007.

Added value of syndromic surveillance for infectious disease surveillance and control

Based on the results of this thesis, we conclude that syndromic surveillance can reveal blind spots of traditional infectious disease surveillance, in particularly by detecting unusual (local) outbreaks independent of diagnoses of specific pathogens, and by improving the monitoring of disease burden and virulence shifts of common pathogens. Syndromic surveillance can also be used for assessing the absence or limited size of unusual disease events, especially during episodes of increased alertness due to epidemics in neighboring countries or during high-profile public events like the Olympics. Generally speaking syndromic surveillance improves the situational awareness of public-health authorities regarding morbidity patterns in the population (chapter 6).

Our results are mostly based on retrospective analysis of syndromic data with high quality and coverage. If real-time collection of syndromic data of high quality and coverage is not feasible, syndromic surveillance for early warning is best avoided, since true outbreaks will then likely be missed and also be swamped by numerous false alarms. For real-time early warning, also sufficient laboratory and epidemiological information is needed, in order to be able to quickly explore possible causes for syndromic signals, and thus recognize signals that need a response.

Compared to the potential costs of delayed response to major infectious disease outbreaks — which can be millions to billions of euros — the costs of maintaining real-time prospective syndromic surveillance are low. Of course some types of outbreaks are more likely to be controlled or stopped after early detection (e.g. SARS) than others (e.g. pandemic influenza), but even if early detection by syndromic surveillance does not lead to control of an outbreak, it can probably still help to scale costly interventions (like quarantine, additional

vaccination etc.) to the size and severity of ongoing outbreaks.

In the Netherlands, prospective surveillance has already started for crude (total) mortality data using the existing mortality registry. Real-time data collection is currently also being implemented for the GP registry. It should be further investigated whether syndromic surveillance can be embedded into (future) real-time data infrastructures, such as the Dutch national health-information-exchange system (EPD). This would possibly allow the inclusion of hospital and other data sources at limited costs in (future) syndromic surveillance systems. In the mean time, periodic updates of retrospective syndromic data sources should be obtained and analyzed to track ongoing infectious disease activity as well as newly emerging disease threats.

This thesis is aimed to contribute to realistic implementation of real-time syndromic surveillance for an improved emerging infectious disease surveillance and control. It illustrates how syndromic data can be used complementary to laboratory data to reveal infectious disease dynamics and outbreaks that otherwise would remain undetected.

Samenvatting

Sinds de SARS epidemie in 2003, de antrax aanslagen in 2001 en de voortdurende dreiging van nieuwe infectieziekte-uitbraken, is vroege detectie van nieuw opduikende infectieziekte-uitbraken een prioriteit geworden in de publieke gezondheidszorg. Maar vroege detectie van uitbraken is gecompliceerd, aangezien patiënten met gevaarlijke infecties als SARS of een nieuw pandemisch influenzavirus zich kunnen presenteren met symptomen typisch voor "normale griep", zoals koorts, vermoeidheid, een beetje misselijk en wat hoesten. Dagelijks melden zich vele patiënten met dergelijke symptomen bij hun (huis)arts, en bij een uitbraak door ongewone of onbekende ziekteverwekkers zal een individuele arts hoogstens enkele extra patiënt(en) zien. Daarom kan niet verwacht worden dat clinici deze ongewone ziektegevallen direct herkennen. Omdat daarnaast vaak ook geen laboratoriumdiagnose wordt gesteld bij zulke gebruikelijke symptomen, kunnen ongewone uitbraken gemist worden. Syndroomsurveillance zou zulke blinde vlekken van de gangbare surveillance – die gebaseerd is op oplettende clinici en laboratoriumsurveillance – kunnen opvangen. Syndroomsurveillance signaleert namelijk verdachte toenames in het aantal patiënten met niet-specifieke klinische symptomen en syndromen zoals "pneumonie door onbekend pathogeen", of toenames in andere niet-diagnostische data zoals het aantal prescripties voor anti-hoest medicatie. Om die reden hebben diverse landen de laatste jaren realtime syndroomsurveillance systemen geïmplementeerd die gebruik maken van pre- of niet-diagnostische data zoals absenteïsme, medische telefonische hulplijnconsulten, spoedeisende eerste hulpconsulten, internet-zoektermen en mediaberichten. Tegelijkertijd wordt echter de effectiviteit van realtime syndroomsurveillance betwijfeld, vooral vanwege mogelijke valse alarmsignalen. Dit proefschrift onderzoekt daarom de toegevoegde waarde van syndroomsurveillance ten opzichte van de gangbare surveillance wat betreft de vroege (realtime) detectie van ongewone infectieziekte-uitbraken en het volgen van de impact van gangbare ziekteverwekkers onder de Nederlandse bevolking.

Voor het onderzoek is allereerst retrospectieve data verzameld uit bestaande Nederlandse medische registraties die mogelijk geschikt zijn voor syndroomsurveillance. Deze registraties bevatten gegevens over werkabsenteïsme, huisartsconsulten, medicatie prescripties, ziekenhuisopnames, laboratoriumtestaanvragen en sterfte. Vervolgens zijn respiratoire, gastro-enteritis en neurologische syndromen gedefinieerd voor deze registraties (hoofdstuk 1). Van deze syndromen kan verwacht worden dat ze de activiteit van zowel gangbare ziekteverwekkers als van ongewone gevaarlijke ziekteverwekkers reflecteren. Dit impliceert dat patiënten met deze ongewone gevaarlijke ziekteverwekkers juist moeilijk te herkennen zijn door clinici, waardoor syndroomsurveillance van extra waarde kan zijn. Een ander voordeel is dat voor deze syndromen gangbare ziekteverwekkers als testvoorbeeld gebruikt kunnen worden. Daarom zijn tijdserie-regressiemodellen gebruikt om te bepalen in hoeverre gangbare ziekteverwekkers bijdragen aan deze syndromen. Deze verklaren het aantal syndroompatiënten per week aan de hand van de aantallen gedetecteerde ziekteverwekkers uit de gangbare laboratoriumsurveillance. Daarna is onderzocht of locale uitbraken zouden kunnen worden gedetecteerd in realtime syndroomsurveillance-data. Dit is gedaan door met behulp van "space-time scan statistics" historische uitbraken te detecteren in de syndroomdata uit de ziekenhuisregistratie. Hieronder worden de belangrijkste onderzoeksresultaten samengevat.

Syndroomsurveillance-data uit medische registraties reflecteert activiteit van ziekteverwekkers.
Hoofdstuk 2 en 6 laten zien dat respiratoire, gastro-enteritis en neurologische syndroom-surveillance-data de gangbare activiteit van ziekteverwekkers reflecteert, en daarom hoogstwaarschijnlijk ook ongewone infectieziekte-activiteit kan reflecteren. Vooral respiratoire syndromen geven een goede reflectie van gangbare activiteit van ziekteverwekkers. In de regressiemodellen kon 68-86% van de wekelijkse syndroomvariaties worden verklaard door wekelijkse aantallen gangbare ziekteverwekkers uit de laboratoriumsurveillance, waarbij influenza en RSV de meeste variatie verklaarden. Voor de gastro-enteritis syndromen werd 29-85% van de syndroomvariatie verklaard door gangbare activiteit van ziekteverwekkers (rotavirus, *Shigella, Campylobacter* en *Salmonella*), en voor een viraal neurologisch syndroom 62% (enterovirus). Verder liepen sommige syndromen tot 5 weken voor op de laboratoriumsurveillance en ook reflecteerden sommige syndromen activiteit van ziekteverwekkers die niet of incompleet werd gedetecteerd door laboratoriumsurveillance.

Verschuivingen in ziektelast en virulentie kunnen worden gevolgd door middel van syndroomsurveillance
In hoofdstuk 3 zijn voor influenza met behulp van tijdserie-regressiemodellen voorheen onbekende jaarlijkse verschuivingen in het aantal ziekenhuisopnames en sterfgevallen ontdekt, gerelateerd aan het aantal gevallen van influenza-achtig-ziektebeeld (IAZ). Deze verschuivingen vallen samen met verschuivingen in antigeniciteit van circulerende influenza virussen. Gecombineerde analyse van syndroom- en ziektespecifieke data maakt het dus mogelijk om de impact van gangbare ziekteverwekkers op de volksgezondheid beter te kwantificeren. Een ander voorbeeld hiervan is de detectie van een voorheen onbekend zware ziektelast van norovirus infecties: een duidelijke associatie is gevonden met milde tot ernstige morbiditeit en zelfs sterfte onder ouderen (hoofdstuk 6). Dat laatste werd eerder wel vermoed, maar kon niet worden bevestigd door de gangbare surveillance.

Nieuwe infectieziekte-uitbraken kunnen tijdig worden gedetecteerd met syndroomsurveillance
De detectie van plots opduikende uitbraken zou kunnen worden versneld indien locale syndroomtoennames - wanneer ze nog te klein zijn om op nationaal niveau te detecteren - al zouden kunnen worden opgespoord. In hoofstuk 4 is dit getest voor bekende historische (zgn. "gouden standaard") uitbraken, en zijn door middel van "space-time scan statistics" de legionella uitbraken in 1999 en 2006 in ziekenhuis syndroomsurveillance-data gedetecteerd. Deze analyse resulteerde in gemiddeld vijf gedetecteerde syndroomclusters per jaar. Dit is een bescheiden totaal aantal alarmsignalen in de tijd en dat duidt erop dat syndroomsurveillance ook locale uitbraken door ongewone of onbekende ziekteverwekkers kan detecteren zonder overspoeld te worden door valse alarmsignalen. Met dezelfde benadering zijn in hoofdstuk 5 ook voorheen onbekende ziekteclusters gedetecteerd, waarvan het aannemelijk is dat ze door Q-koorts zijn veroorzaakt. Dit illustreert dat syndroomsurveillance in sommige situaties uitbraken kan detecteren die anders gemist worden. Realtime detectie en het nader onderzoeken van deze Q-koorts verdachte clusters zou wellicht geleid hebben tot signalering van toegenomen Q-koorts activiteit vòòr de detectie die nu in 2007 plaats heeft gevonden.

Toegevoegde waarde van syndroomsurveillance voor infectieziektesurveillance

Op basis van de resultaten in dit proefschrift wordt geconcludeerd dat syndroomsurveillance blinde vlekken van gangbare infectieziektesurveillance kan compenseren. De toegevoegde waarde van syndroomsurveillance zit in het detecteren van ongewone uitbraken onafhankelijk van laboratoriumdetectie van de ziekteverwekker en in het signaleren van verschuivingen in virulentie en ziektelast van gangbare ziekteverwekkers. Daarnaast kan syndroomsurveillance ook gebruikt worden om de afwezigheid of beperkte omvang van ongewone ziekte-uitbraken in de bevolking vast te stellen, vooral tijdens periodes van toegenomen alertheid, bijvoorbeeld tijdens epidemieën in naburige landen of tijdens grote publieke evenementen zoals de Olympische Spelen. Samenvattend verbetert syndroomsurveillance het inzicht van volksgezondheidsautoriteiten wat betreft het vóórkomen van ziektes onder de bevolking.

De resultaten in dit proefschrift zijn vooral gebaseerd op retrospectieve analyses van syndroomsurveillance-data met weinig data-artefacten en hoge dekkingsgraad. Als realtime collectie van dergelijke syndroomsurveillance-data niet haalbaar is, is het beter syndroomsurveillance niet in te zetten voor vroege (realtime) signalering van ziekte-uitbraken. Dit aangezien echte uitbraken dan vaak gemist zullen worden en tevens vele valse alarmsignalen gegenereerd kunnen worden. Verder zijn voor vroege signalering ook voldoende laboratorium- en epidemiologische gegevens nodig om snel mogelijke oorzaken voor syndroomsignalen te kunnen exploreren, waardoor alarmsignalen die een respons nodig hebben beter kunnen worden onderscheiden.
Vergeleken met de mogelijke kosten van een late respons op grote infectieziekte-uitbraken - die miljoenen of zelfs miljarden euro's kunnen bedragen - zijn de kosten voor het onderhouden van realtime syndroomsurveillance laag. Natuurlijk kunnen sommige ziekte-uitbraken makkelijker worden beheerst of gestopt (bv. SARS) dan anderen (bv. pandemische influenza). Maar zelfs als vroege detectie door syndroomsurveillance niet leidt tot beheersing van een uitbraak, dan nog kan syndroomsurveillance van nut zijn om vaak dure interventiemaatregelen, zoals quarantaine en aanvullende vaccinatie, beter af te stemmen op de omvang en ernst van de uitbraak.

Bij de start van dit onderzoek was in Nederland nog geen sprake van realtime syndroomsurveillance. Inmiddels is begonnen met een wekelijkse sterftesurveillance welke gebruik maakt van data uit de bestaande mortaliteitsregistratie. Realtime datacollectie wordt momenteel ook geïmplementeerd voor de in dit proefschrift genoemde landelijke huisartsconsult-registratie (hoofdstuk 2 en 6). Het zou verder onderzocht moeten worden of collectie van syndroomsurveillance-data kan worden ingebouwd in (toekomstige) realtime medische registraties, zoals het elektronisch patiëntendossier (EPD). Wellicht kunnen ziekenhuisdata en andere datasoorten zo tegen beperkte meerkosten beschikbaar worden gemaakt voor syndroomsurveillance. Tot het zover is kan gebruik gemaakt worden van retrospectieve syndroomsurveillance-data om, weliswaar met een vertraging, de impact van infectieziekten onder de Nederlandse bevolking te volgen en nieuw opduikende infectieziektedreigingen te signaleren.

Dit proefschrift beoogt bij te dragen aan een realistisch gebruik van realtime syndroomsurveillance voor een betere signalering van nieuw opduikende infectieziekteproblemen. Het illustreert hoe syndroomsurveillance,

complementair aan gangbare surveillance, nieuwe infectieziekte-uitbraken en impact van gangbare ziekteverwekkers kan detecteren die anders gemist zouden worden.

Dankwoord/Acknowledgements

Het dankwoord schijnt de meest gelezen tekst van een proefschrift te zijn, terwijl het juist niet over de inhoud van het promotie-onderzoek gaat. Daarom heb ik even overwogen geen dankwoord te schrijven, maar dat ging me te ver. Veel mensen hebben mij namelijk de afgelopen jaren op allerlei manieren gesteund en die wil ik niet ongenoemd laten.

Allereerst wil ik graag mijn promotoren Marion Koopmans en Nico Nagelkerke, en mijn co-promotor Wilfrid van Pelt bedanken. Marion, jij maakte mogelijk dat het onderzoek naar syndroomsurveillance kon worden opgezet en hebt me vervolgens gestimuleerd om daarop te promoveren. Ik heb veel gehad aan je ideeën en kritische opbouwende commentaar en ik denk dan ook met genoegen terug aan de werkbesprekingen op het LIS de afgelopen jaren. Nico, net voordat je wegging bij het RIVM was je betrokken bij de opzet van het syndroomsurveillance onderzoek. Ik ben heel blij dat je ook in de jaren daarna bereid was mij vanaf grote afstand bij te staan met vaak cruciaal advies. Het was een mooie ervaring om op 1 dag heen en weer naar Genève te vliegen en daar samen met jou op het vliegveld een probleem met de analyses op te lossen. Wilfrid, jouw deur stond altijd open en daar heb ik dan ook veelvuldig gebruik van gemaakt. Jij droeg veel van de ideeën voor de analyses aan en had altijd creatieve oplossingen als we op problemen stuitten. En als ik terugdenk aan ons bezoek aan het syndroomsurveillance congres in New York schiet ik nog steeds in de lach.

Ik ben vereerd door de samenstelling van mijn promotiecommissie en wil alle leden danken voor hun bereidheid plaats te nemen in de commissie.

Dit promotie onderzoek was alleen mogelijk door de medewerking en het verstrekken van data door het Nederlands instituut voor onderzoek van de gezondheidszorg (Nivel), de Stichting Farmaceutische Kengetallen (SFK), Dutch Hospital Data (DHD) en het Centraal Bureau voor de Statistiek (CBS). Deze organisaties en de betrokken medewerkers wil ik dan ook hartelijk danken. Ook wil ik alle co-auteurs hartelijk bedanken vanwege hun onmisbare bijdrage aan dit proefschrift.

Mijn promotie onderzoek heb ik uitgevoerd binnen mijn baan als onderzoeker op de afdeling Epidemiologie en Surveillance van het RIVM-Centrum voor Infectieziektebestrijding. Graag wil ik de achtereenvolgende afdelingshoofden Marina Conyn-van Spaendonck en Marianne van der Sande bedanken voor het creëren van de randvoorwaarden voor dit promotieonderzoek. Dank ook aan de hele EPI-afdeling voor de prettige werksfeer tijdens gewone werkdagen en gezelligheid tijdens lab-uitjes of heidesessies. Daarnaast wil ik nog een aantal collega's speciaal noemen. Allereerst veel dank aan mijn directe collega Liselotte van Asten. Jij kwam kort na de start van het syndroomsurveillance onderzoek ons team versterken. Ik kan me veel enthousiaste brainstormsessies herinneren, bijvoorbeeld tijdens de congressen die we samen hebben bezocht. Ook al ging onze samenwerking niet altijd vanzelf, we konden steeds samen weer verder en daar heb ik veel van geleerd. Hans van Vliet, jij bent gedurende het grootste deel van mijn promotietraject mijn directe leidinggevende geweest, en dat was precies wat ik nodig had want je bent een goede coach (en dat is zeker niet cynisch bedoeld). Mirjam Kretzschmar, het laatste jaar heb ik van jou als nieuwe leidinggevende niet alleen de ruimte gekregen om mijn promotie af te ronden, maar je hebt me ook met raad en daad bijgestaan. Liesbeth Mollema, jij was/bent een heel fijne kamergenoot! In 2005 gaf jij me een lift vanaf de bushalte in Wageningen naar het RIVM en vervolgens bleek je mijn nieuwe kamergenoot te zijn. Sinds die tijd heb ik heel wat met je afgelachen, maar kon ik bij jou zonodig ook mijn verhaal kwijt, en gaf je nuchtere en bruikbare feedback. Heel fijn dat je mijn paranimf wilt zijn. Ook niet onbelangrijk voor nieuwe inspiratie was de "wandelclub". Bedankt Jan, Els, Frederika, Alies, Jelle, Jan-Dirk en de vele anderen met wie ik in de loop der jaren het Houdringe lunchrondje heb gelopen.

Special thanks to the international colleagues of the ISDS (International Society for Disease Surveillance) and the European syndromic surveillance network. You have greatly contributed to this thesis by the inspiring presentations and discussions at the annual syndromic surveillance conferences in the USA and the European syndromic surveillance workshops.

Afleiding en gezelligheid gaven me de afgelopen jaren nieuwe energie voor het promotiewerk. Die afleiding en gezelligheid was vaak afkomstig van vrienden en die wil ik daarvoor dan ook van harte bedanken. Erik, Willem, Huub & Cindy, Kees & Geke, Steven, Jan & Ingrid, Bas & Anne-Margreet, Bas & Nienke, Auke-Jan & Lianne, en alle andere Vadje mannen bedankt, volgend jaar ga ik wel mee op vakantie! Ook veel dank voor de gezellige bestuursetentjes aan Anne, Ids & Francine, Kees & Emmeke, Marjakke & Lennert, Marleen, Martin & Anke en Nelleke & KP, binnenkort weer een keer in Zeist! Joris & Mariska, Rik & Yvette en Robert & Leona bedankt voor de jarenlange vriendschap en leuke nostalgische weekendjes, wat mij betreft gaan we een lang weekend naar Noorwegen binnenkort. Robbert, jij hebt waarschijnlijk aan de wieg gestaan van deze promotie door onze competitie bij meneer Möhlmann. Bedankt daarvoor en leuk dat jij en Sasja in deze levensfase gelijk met ons blijken op te trekken, laten we snel weer met de "families" afspreken. Arnold & Annemarie, "beter een goede buur dan een verre vriend" wordt wel eens gezegd, en jullie zijn ook nog goede vrienden geworden. Nu kunnen jullie eindelijk lezen waar "die discussie" nou eigenlijk over ging.

Papa, mama maakt dit helaas niet meer mee, maar ik wil jullie hier samen bedanken voor de interesse die jullie altijd toonden bij alles wat ik ondernam, en ook de steun die jullie gaven als dat nodig was. Papa & Maudy, Karel & Judith, Linda, Koen en Kasper en de rest van de familie bedankt voor alle gezelligheid en afleiding de afgelopen jaren. Karel, heel fijn dat jij mijn paranimf wilt zijn. Ook mijn schoonfamilie Aad & Ineke, Marnix & Yvonne, Joep & Marieke en nichtje en neefjes dank voor de gezellige familiebijeenkomsten op de boerderij en de praktijklessen in nuchterheid voor het dagelijks leven.

En dan volgt nu het thuisfront. Lieve Evelien, onze basis samen met Lars en sinds kort Maarten geeft me enorm veel energie en geluk. Dat heeft me bovenal gesteund bij het afronden van dit proefschrift. Daarnaast kon ik natuurlijk niet achterblijven na de bewonderenswaardige manier waarop jij twee jaar geleden promoveerde net na de geboorte van Lars, dat was een mooie extra stok achter de deur.

Met voldoening zet ik nu de laatste.

Kees

Curriculum Vitae

Kees van den Wijngaard werd geboren op 5 februari 1976 te Sittard. Na het behalen van zijn vwo diploma aan het Gymnasium Beekvliet in St. Michielsgestel, studeerde hij dierwetenschappen aan de Wageningen Universiteit. Tijdens deze studie deed hij een eerste afstudeerstage voor Holland Genetics om de verspreiding van BHV1-virus onder fokstieren te modelleren. Een tweede afstudeerstage deed hij aan de Universiteit van Montreal, Canada. Daar heeft hij onderzoek gedaan naar risicofactoren voor embryonale sterfte bij melkvee in Quebec en naar de kwaliteitsparameters van drachtigheidstesten voor melkvee. Voor een derde afstudeerstage, via de vakgroep Agrarische Bedrijfseconomie, maakte hij voor Interpolis een economische analyse van mestafzetconstructies en daarmee gepaard gaande financiële risico's voor varkensbedrijven in Zuid-Nederland. In 2001 studeerde hij af en begon als plaatsvervangend docent aan Wageningen Universiteit bij de vakgroep Kwantitatieve Veterinaire Epidemiologie. In 2002 stapte hij over naar de Gezondheidsdienst voor Dieren in Deventer waar hij heeft gewerkt aan internetcommunicatie richting veehouders over dierziektepreventie.

In januari 2003 trad hij in dienst als onderzoeker bij de afdeling Epidemiologie en Surveillance van het Centrum voor Infectieziektebestrijding van het RIVM. Vanaf dat moment hield hij zich vooral bezig met humane infectieziekte-epidemiologie. Daarbij heeft hij zich gericht op signaleringsmethoden voor zogenaamde "emerging" infectieziekte-uitbraken. Dit heeft geleid tot het syndroomsurveillance onderzoeksproject dat hij samen met Dr.ir. L. van Asten heeft opgezet, begeleid door Dr. W. van Pelt, Prof.dr. N.J.D. Nagelkerke en Prof. dr. M.P.G. Koopmans. Dit onderzoeksproject heeft geresulteerd in een aantal publicaties die de basis vormden voor dit proefschrift. In 2007 organiseerde hij samen met zijn collega Dr.ir. L. van Asten een Europese syndroomsurveillance workshop in Bilthoven en was hij "Program Committee Member" van het jaarlijkse syndroomsurveillance congres in de VS.

Erasmus MC
Universitair Medisch Centrum Rotterdam

PhD Portfolio

Summary of PhD training and teaching

Name PhD student: ir. C.C. van den Wijngaard Erasmus MC Department: Virology	PhD period: 2003-2010 Promotor(s): Prof.dr. M. P. G. Koopmans Prof.dr. N. J. D. Nagelkerke Supervisor: Dr. W. van Pelt

1. PhD training

	Year	Workload (Hours/ ECTS)	
		Hours	ECTS
General courses			
- Modern Statistical methods (NIHES)	2005		4.3
- NIHES, Principles of Epidemiologic Data-analysis (EP15)	2007		1.4
- Scientific writing in English	2009	56	
Specific courses (e.g. Research school, Medical Training)			
- Several SAS courses RIVM	2003-2007	32	
- Statistical course RIVM	2008	24	
- Erasmus Summer Programme courses	2006	40	
- Epiet and ECDC courses	2003-2008	200	
Seminars and workshops			
- (co-)Organizing Symposium Advances in Disease Surveillance, Bilthoven 22 juni 2007	2007	28	
- Workshop on Syndromic surveillance in Europe: the practice of early detection of health events	2008	14	
Oral presentations			
- Syndromic Surveillance (epiet Alumni meeting, Lyon)	2004	16	
- Syndromic approach in The Netherlands to the West Nile virus threat (2004 Syndromic Surveillance Conference, Boston)	2004	20	
- Syndromic approach to WNV threat in the NL (VEEC)	2005	4	
- Syndromic Surveillance for a Large Respiratory Disease Outbreak by Legionella in the Netherlands (2006 ISDS conference, Baltimore)	2006	20	
- Selection of datasets, evaluation of Syndromic Surveillance potential (Symposium Advances in Disease Surveillance, Bilthoven 22 juni 2007)	2007	20	
- Dutch experience in evaluating syndromic surveillance (SS): a retrospective study (ESCAIDE, Stockholm)	2007	20	
- Syndromic surveillance: space-time signal interpretation (S-gem workshop on detection algorithms, Stockholm)	2009	20	
- Q fever outbreaks: Syndromic approach for detection of hidden clusters (ISDS conference, Miami).	2009	20	
(Inter)national conferences			
- Annual syndromic surveillance conference of International Society for Disease Surveillance (ISDS) (except 2005)	2003-2009	168	
- European Scientific Conference on Applied Infectious Disease Epidemiology' (ESCAIDE)	2007	28	

2. Teaching

	Year	Workload (Hours/ ECTS)	
		Hours	ECTS
Lecturing			
- Syndromic surveillance for early warning of bioterrorism and other unusual infectious disease events (Master students Nijmegen University)	2005	28	
- Syndromic surveillance: rationale, methods and applications. (Master students Nijmegen University)	2007	14	
- Innovations in syndromic surveillance (Research master 'Infection & Immunity' Erasmus MC Graduate School)	2010	14	
Supervising practicals and excursions, Tutoring			
- Assistant during the course "Surveillance in de infectieziektenbestrijding" (Netherlands School of Public & Occupational Health)	2004	8	
TOTAL		794	5.7