

RESEARCH ARTICLE

Open Access

Automated data extraction from general practice records in an Australian setting: Trends in influenza-like illness in sentinel general practices and emergency departments

Gösta TH Liljeqvist^{1,2*}, Michael Staff³, Michele Puech³, Hans Blom⁴ and Siranda Torvaldsen²

Abstract

Background: Influenza intelligence in New South Wales (NSW), Australia is derived mainly from emergency department (ED) presentations and hospital and intensive care admissions, which represent only a portion of influenza-like illness (ILI) in the population. A substantial amount of the remaining data lies hidden in general practice (GP) records. Previous attempts in Australia to gather ILI data from GPs have given them extra work. We explored the possibility of applying automated data extraction from GP records in sentinel surveillance in an Australian setting.

The two research questions asked in designing the study were: Can syndromic ILI data be extracted automatically from routine GP data? How do ILI trends in sentinel general practice compare with ILI trends in EDs?

Methods: We adapted a software program already capable of automated data extraction to identify records of patients with ILI in routine electronic GP records in two of the most commonly used commercial programs. This tool was applied in sentinel sites to gather retrospective data for May-October 2007-2009 and in real-time for the same interval in 2010. The data were compared with that provided by the Public Health Real-time Emergency Department Surveillance System (PHREDSS) and with ED data for the same periods.

Results: The GP surveillance tool identified seasonal trends in ILI both retrospectively and in near real-time. The curve of seasonal ILI was more responsive and less volatile than that of PHREDSS on a local area level. The number of weekly ILI presentations ranged from 8 to 128 at GP sites and from 0 to 18 in EDs in non-pandemic years.

Conclusion: Automated data extraction from routine GP records offers a means to gather data without introducing any additional work for the practitioner. Adding this method to current surveillance programs will enhance their ability to monitor ILI and to detect early warning signals of new ILI events.

Background

Influenza poses a significant threat to public health, not only in the form of pandemics but also as seasonal epidemics. Each year, influenza causes 3-5 million cases of severe illness, and up to 500 000 deaths globally [1]. Surveillance of influenza-like illness (ILI) enables detection of the onset of influenza seasons and can provide early warning of new events. The National Health

System in the United Kingdom has gained much from automated ILI data extraction in the QFlu program [2], and Brabazon *et al.* [3] obtained promising results with automated electronic data extraction from general practitioner (GP) records in Ireland.

In Australia, GPs play a key role in managing influenza and are often the first point of contact for patients with ILI [4]. Previous GP surveillance for ILI in Australia has required some form of additional work by GPs, either filling in reports or actively following surveillance-guided medical reporting using coded sections of medical records programs [5]. In NSW, notifications of

* Correspondence: Gosta.Liljeqvist@doh.health.nsw.gov.au

¹NSW Public Health Officer Training Program, New South Wales Department of Health, Sydney, New South Wales, Australia

Full list of author information is available at the end of the article

laboratory-confirmed cases [6] and the Public Health Real-time Emergency Department Surveillance System (PHREDSS) [7] are the main data sources for influenza surveillance. While laboratory notification data include the outcomes of GP consultations, they provide notifications only of positive laboratory samples, severely limiting their value. Increasing numbers of GPs use software packages for keeping medical records [8], which has improved the uniformity and the quality of the records [9]. This may facilitate timely, comprehensive GP surveillance, which would enhance our ability to achieve early detection of new trends and identify and monitor clusters of cases [10].

We tested electronic automated data extraction of ILI data from routine GP records, assessed compatibility of the adapted software with GP office networks and investigated the timeliness of data and the sensitivity and specificity of automated ILI syndromic surveillance. We also compared local GP-reported ILI patterns with those seen at local emergency departments (EDs).

The aim of this study was to assess if syndromic ILI data can be extracted automatically from routine GP data by adapting an existing data extraction tool in Australia and to compare ILI data from general practice with ED data at a local level.

Methods

Ethical approval

Ethical approval for this study was provided by the Harbour Research Ethics Committee of Northern Sydney Central Coast Health.

Developing a new data extraction instrument

The Canning Division of General Practice in Western Australia has developed a software application capable of automated data extraction from GP records [11]. This tool is already in use for collection of data on chronic diseases, such as diabetes, but in its current configuration cannot be used for free-text searches. Discussions with local GP networks and potential participating GPs indicated that the Canning Tool was known and well accepted. Of 110 divisions of general practice in Australia, 85 use the Canning Data Extraction tool [11]. This information and the compatibility with GP medical records programs led us to select the Canning Data Extraction Tool for developing our new application.

On the platform of the existing Canning Tool, we developed a software application capable of automated data extraction from routine GP records. Free-text and coded information search combinations were used to identify cases that met an adapted case definition of ILI: *Any patient presenting with cough, fever or self-reported fever and fatigue* [12]. Free-text clinical information can be entered into the medical record, and provisional

diagnoses can be recorded either in tick boxes, drop-down menus or free text. Information entered into tick boxes or drop-down menus is stored as codes rather than free text in the medical record database and requires a separate extraction process.

Building on the Canning Data Extraction Tool, we developed the Canning Flu Tool to conduct automated searches and extractions of both coded and free-text fields from two of the most commonly used GP medical record packages used in Australia (Best Practice and Medical Director 3). As the case definition has three distinct elements: respiratory symptoms, fever and fatigue, we programmed the tool to select words that described these elements (Table 1): a case was identified as ILI if at least one word in each of the three categories was present in the record. We also programmed the tool to recognise the expressions 'influenza', 'flu', 'ILI' and 'H1N1', and to use these as positive triggers when they appeared unaccompanied in the record. In order to address the possibility that any of these words was entered as a negative, i.e. no influenza or fever absent, each word was accompanied by a complex structured query language code for exclusion of negating terms.

Testing the Canning Flu Tool

The extraction program of the Canning Flu Tool was refined initially with a mock database and later with a real patient database. This allowed systematic testing of a variety of word combinations for both inclusion and exclusion in the ILI category. It also provided the opportunity to pilot-test the software in a real practice setting to ensure its compatibility with a practice computer network. This process was done with a group of participating GPs, who gave advice on common recording practices, and programmers to adjust and refine the data extraction process. The early version of the tool identified all visits for influenza vaccination as positive for ILI because the word "flu" appeared in these records. This resulted in a large number of false positives. The problem was corrected by adding specific exclusion criteria for the terms used to describe influenza vaccination. Similar processes were required to address the issue of patients visiting their GP to discuss pandemic influenza without actually having any symptoms.

After refinement, the tool's sensitivity and specificity for detecting cases that met the case definition were assessed by comparing its performance against the professional opinion of two public health physicians, who reviewed all the electronic records from two participating practices with Best Practice software during 1 week in August 2008.

Program design and data collection

After refinement of the application, a study was conducted in the northern metropolitan area of Sydney,

Table 1 Canning Flu Tool search triggers

Canning Flu Tool trigger words for free text search		
Case definition: A person presenting with cough, fever $\geq 38^{\circ}\text{C}$ (or self-reported history of fever) and fatigue		
The Canning Flu Tool identified as positive for ILI, any record that contained at least one word from each of the three columns (the search was not case sensitive)		
Respiratory	Fever	Fatigue
Cough	Chill	Fatigue
Dyspnoea	Fever	Lethargy
Respiratory Tract	Feverish	Lethargic
Infection	Febrile	Malaise
RTI	Pyrexia	Prostrate
Shortness of Breath	Pyrexia	Prostration
SOB	Rigor	Tired
Wheeze	Temp	Tiredness
Wheezing	Temps	Unwell
	Temperature ≥ 38	
	Temp ≥ 38	

The terms Flu, Influenza, ILI and H1N1 triggered a positive categorisation without the need for another word
 Any description of influenza vaccination caused a negative categorisation.

which has about 820 000 residents [13], five public EDs, and an estimated 1048 general practitioners in 328 practices [14], most being members of one of three local GP networks.

Eight GP sites participated in the study; in all, they saw up to 7000 patients per week (300 to 2000 patients per week per practice) during the surveillance period. Practices were selected on the basis of their geographical location in order to represent the study area and also to broadly match the catchment areas of the five public EDs in northern Sydney. Our local GP networks informed us that most GPs in the area use electronic records but that some do not use them to their full capacity, which would limit their use for surveillance purposes. We were guided by the GP networks in selecting practices, but this did not impose any limitations on selection of strategic sentinel practices. Our selection criteria were: use of either Best Practice or Medical Director 3, willingness to participate, and at least one practice within the catchment area of each ED.

We collected weekly reports from the beginning of May 2010 to the end of October 2010 in real-time and also collected retrospective data for comparison from all but one practice for the period May to October in 2007-2009 (the excluded practice did not open until 2009). The tool automatically extracted aggregated de-identified data and was installed either on a network server or on a work station in the practice. Data from all five local EDs were included, contributing a little over 2500 presentations per week. The average weekly counts of all ED presentations were 2749 (2007), 2777 (2008), 2972 (2009) and 2979 (2010). Weekly automated reports were delivered by the tool from each GP sentinel site

including the total number of visitors per week and the count of ILI cases per week. Counts were conducted so that an individual could only contribute once to the number of patients who presented in any single week.

The weekly percentage of ILI cases in each practice was calculated automatically by the tool using the count of visits as denominator and the count of ILI cases as numerator.

When measuring sentinel activity, we used the weekly ILI percentage per site as provided in the weekly automated reports and calculated the arithmetic mean of these percentages to obtain a weekly measurement of ILI activity. We favoured this approach over weekly crude counts to ensure that summary measures were not unduly influenced by individual sites. It also made it possible to set a threshold for widespread influenza activity. We applied a 2% threshold in this study, which is consistent with practice elsewhere [15].

Feedback was provided to GPs by giving them access to a password-protected, secure website containing weekly surveillance reports, including both practice-specific and area-wide surveillance data and interpretations. Practice-specific information was coded in the reports so that it was identifiable only to the individual practice.

Comparing emergency department trends with general practice trends

We conducted a descriptive analysis to compare ILI trend curves. The data from both EDs and the sentinel GP sites were displayed in graphs using Microsoft Excel. GP data were temporally compared with data from the five local EDs collected by PHREDSS for the 2007-2009 periods and the 2010 period. PHREDSS extracts provisional diagnoses of ILI manually recorded in ED triage

notes and coded according to ICD-9, ICD-10 or the Systematized Nomenclature of Medicine (SNOMED) for diseases caused by influenza [16]. At the time of this study, all the participating EDs used ICD-9.

In addition, we performed a free-text structured query language search of the ED records using the same search term as that used for GP records. The comparison of EDs and sentinel GP sites was performed retrospectively for May - October 2007, 2008 and 2009.

Results

Sensitivity and specificity of the Canning Flu Tool

Records from two practices for a 1-week period in August 2008 were reviewed (Table 2). The sensitivity of the tool was 96.3% and the specificity 99.7%. The positive predictive value was 76.5%, while the negative predictive value was 99.7%.

Sentinel GP data compared with PHREDSS data

Between 2007 and 2010, clear seasonal peaks can be seen in the GP data, the 2% threshold being exceeded in all four years (Figure 1). The percentage of ILI in the PHREDSS data remained below 1% in all non-pandemic years and did not provide a clear visual signal, while the sentinel GP sites showed clear seasonal peaks in each season monitored. In 2009, the GP peak preceded the ED peak by about 2 weeks; the amplitude was not great enough in the other years to allow a visual comparison.

In the non-pandemic ILI seasons monitored, the number of weekly ILI presentations in the five EDs ranged from 0 to 18 with percentages ranging from 0 to 0.93 in 2007 and from 1 to 13 with percentages from 0 to 0.54 in 2008, while the number in the GP sentinel sites ranged from 8 to 128 per week with percentages from 0.27 to 3.48 in 2007 and from 16 to 108 per week with percentages from 0.78 to 2.77 in 2008. Both the ED and the GP data showed clear increases associated with the pandemic season in 2009. The total count of visits per

week in the five combined EDs ranged from 2441 to 3474 (the highest being in July 2009), with a mean of 2865. The highest total count of visits in the sentinel general practices was 7770 in October 2010. The total number of visits to GPs increased progressively from year to year, with the lowest count in June 2007 at 2432 (one of our sentinel practices opened in 2009).

Free text extraction from ED data

A free-text search of ED records resulted in a seasonal curve that more closely matched the GP ILI curve than that provided by ED provisional diagnoses (as in PHREDSS). The percentage of syndromic ILI in EDs exceeded that at the GP sentinel sites when a free-text search was used (Figure 2).

Discussion

Our results indicate that the Canning Flu Tool would be valuable for monitoring ILI activity in general practice and that it provides a sensitive signal of seasonal patterns of influenza like illness on a near real-time basis. Unlike other systems in Australia, the Canning Flu Tool can collect data from GP records without imposing any additional tasks. The tool produced a more robust signal than PHREDSS and is likely to detect increased ILI activity more reliably, chiefly because the counts are higher and less volatile, while cases of ILI as per the case definition are still detected with high sensitivity and specificity. The sensitivity of PHREDSS was enhanced by application of a free-text search, which detected more cases than provisional diagnostic coding and resulted in a temporal pattern very similar to that from GP data.

Adding the Canning Flu tool to NSW and Australian surveillance practice could improve the detection of seasonal onset of influenza like illness and hence assist GPs to prepare for the inevitable influenza season each year, and allow timely community messages and increased infection control measures, which are vital parts of both

Table 2 Canning Flu Tool sensitivity and specificity table

Sensitivity and specificity of the Canning Flu Tool: Test against the clinical opinions of two Public Health Physicians using records from two practices (n 2518) over two weeks in August 2008			
		Gold Standard (Public Health Physician's clinical opinion)	
CFT	ILI present	ILI not present	Total
+	26	8	34
-	1	2483	2484
Total	27	2491	2518

Legend. Canning Flu Tool (CFT), ILI (Influenza Like Illness)

Sensitivity 96.3%

Specificity 99.7%

Positive predictive value 76.5%

Negative predictive value 99.7%

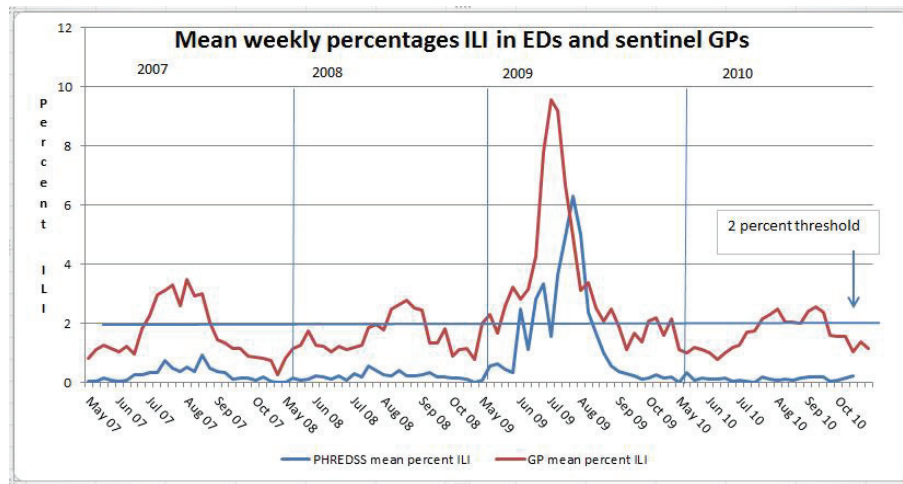


Figure 1 Weekly mean percentages of ILI at GP sentinel sites and in five local EDs. PHREDSS mean percentage ILI, GP mean percentage ILI.

pandemic and seasonal preparedness at frontline services such as GPs and EDs [17]. Furthermore, this use of electronic data opens up a previously largely unmonitored portion of ILI activity and hence provides valuable epidemiological knowledge, resulting in increased ability for early detection of new trends. This will further inform and enhance public health responses.

By design, the Canning Flu Tool identifies ILI rather than influenza specifically. It provides syndromic information and seasonal trend observations and should hence be used in combination with and as a complement to laboratory testing and other surveillance methods. Furthermore, the study was conducted in a

metropolitan area with relatively high socioeconomic status, and we cannot assume that the findings are applicable to the whole population of NSW or Australia.

One of the two medical record software packages accessed by the Canning Flu Tool allowed free-text searching of the progress notes field, while the other package encrypts the progress notes field. The sensitivity of the tool might be lower in practices in which the latter software package is used, and this should be considered when recruiting GPs to a surveillance system based on electronic medical records.

For further analysis of trends, a formal time series analysis would be ideal, but this has some pitfalls, as described

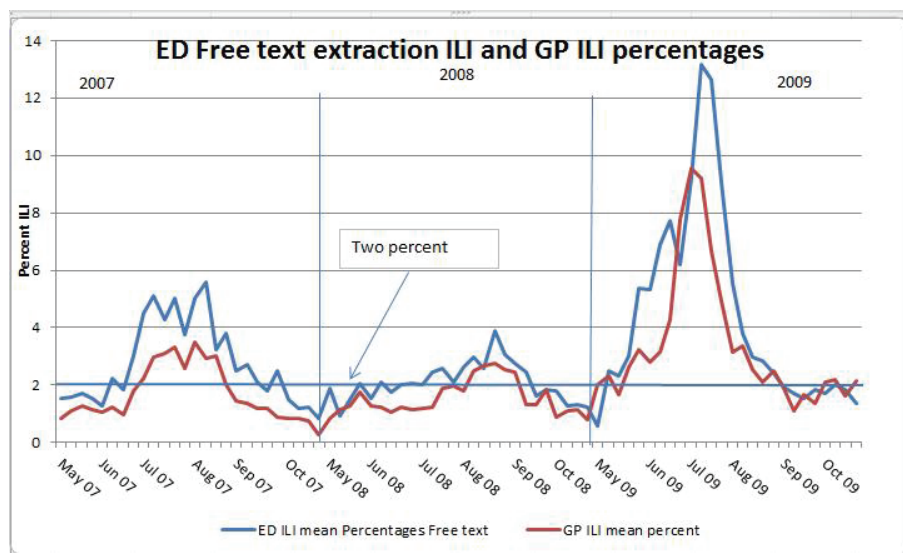


Figure 2 Free-text search for ILI in EDs compared with GP sentinel sites (mean percentage). ILI weekly percentages in EDs and sentinel GPs with the same free-text extraction code used for both data sets.

by Zheng et al. [18], and requires rich data, i.e. daily counts, which was beyond the scope of this study.

The Canning Flu Tool can provide timely, locally relevant information for GPs to enhance infection control measures within their practices, give targeted messages to their patients and have greater collaboration and communication with public health authorities. Future studies are recommended to develop additional syndromic applications for automated data extraction from GP records. The Canning Flu Tool could be expanded to other relevant syndromes, either for locally specific programs or for state- or nationwide surveillance. This might result in a significant improvement in surveillance not only of infectious diseases but potentially also of chronic conditions.

As electronic records are being used by increasing numbers of GPs in Australia, and automated data extraction from electronic records has proven to be useful in the United Kingdom and Ireland [2,3], we consider that this could become an efficient surveillance method in the Australian setting, where general practices are mostly private.

Conclusion

This study demonstrates that the Canning Flu Tool performs well and has the potential to be useful within the Australian general practice setting. Automated data extraction from routine GP records offers a means to gather data without introducing any additional work for the practitioner. Adding this method to current surveillance programs will enhance their ability to monitor ILI and to detect early warning signals of new ILI events.

Acknowledgements

We would like to thank Mr Ian Peters of the Canning Division of General Practice for his significant input into this study by programming the Canning Flu Tool. We would also like to thank Mr David Muscatello (NSW Health PHREDSS), Ms Robin Gilmour (NSW Health Communicable Disease Branch), Mr Terry Black (Northern Sydney Central Coast Area Health Service), Manly Warringah Division of General Practice, GP Network Northside, Northern Sydney General Practice Network and GP NSW, without whom this project would not have been possible. Professor Elisabeth Heseltine provided instrumental editorial support.

Author details

¹NSW Public Health Officer Training Program, New South Wales Department of Health, Sydney, New South Wales, Australia. ²School of Public Health and Community Medicine, University of New South Wales, Sydney, New South Wales, Australia. ³Northern Sydney Central Coast Area Health Service, Hornsby, New South Wales, Australia. ⁴Vale Medical Centre, Brookvale, New South Wales, Australia.

Authors' contributions

GL was the lead author and project manager; he collected and analysed data and provided interpretation. MS provided study design support and guidance as well as data analysis support, assessment of the Canning Flu Tool (sensitivity and specificity, data analysis) and editorial support. MP participated in assessment of the Canning Flu Tool (sensitivity and specificity, data analysis) and gave editorial support. HB Tested the Canning Flu Tool in a practice setting, provided input on GP recording practices for

the development of search triggers and gave editorial support. ST contributed to the analysis and interpretation of data, critically revised the manuscript for intellectual content and gave editorial advice. All the authors approved the final version of the manuscript.

Competing interests

None. The Canning Division of General Practice was consulted and remunerated for developing their tool to the specifications of the study. The chief project manager was employed by the NSW Department of Health with an affiliation to the University of New South Wales.

Received: 15 February 2011 Accepted: 6 June 2011

Published: 6 June 2011

References

1. World Health Organization: **Influenza (seasonal)**. 2009 [http://www.who.int/mediacentre/factsheets/fs211/en].
2. Hippisley-Cox J, Smith S, Smith GE, Porter A, Heaps M, Holland R, Fenty J, Harcourt S, George R, Charlett A, Pebody RG, Painter M: **QFLU: new influenza monitoring in UK primary care to support pandemic influenza planning**. *Euro Surveill* 2006, **11**(25):2980[http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2980].
3. Brabazon ED, Carton MW, Murray C, Hederman L, Bedford D: **General practice out-of-hours service in Ireland provides a new source of syndromic surveillance data on influenza**. *Euro Surveill* 2010, **15**(31):19632[http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19632].
4. Dalton DB, Pearce R, Litt JCB: **ASPREN surveillance system for influenza-like illness. A comparison with FluTracking and the Notifiable Diseases Surveillance System**. *Aust Fam Physician* 2009, **38**(11):932-936[http://www.racgp.org.au/afp/200911/35801].
5. Clothier HJ, Fielding JE, Kelly HA: **An evaluation of the Australian Sentinel Practice Research Network (ASPREN) surveillance for influenza-like illness**. *Commun Dis Intell* 2009, **29**(3):231-247.
6. NSW Health: **Guides and caveats for interpreting infectious diseases data (NSW data)**. [http://www.health.nsw.gov.au/data/diseases/caveats.asp].
7. NSW Health: **Public Health Real-time Emergency Department Surveillance System (PHREDSS) Public Health Unit Response**. 2010 [http://www.health.nsw.gov.au/policies/gl/2010/GL2010_009.html].
8. Australian Government: **Australian Government Productivity Commission Report on Government Services**. Primary and Community Health; [http://www.pc.gov.au/gsp/reports/rogs/2010], Chapter 11.
9. McInnes DK, Saltman DC, Kidd MR: **General practitioners' use of computers for prescribing and electronic health records: results from a national survey**. *Med J Aust* 2006, **185**(2):88-91[http://www.mja.com.au/public/issues/185_02_170706/mci10476_fm.html].
10. Robertson C, Nelson TA, MacNab YC, Lawson AB: **Review of methods for space-time disease surveillance**. *Spat Spat-Temp Epidemiol* 2010, **1**(2-3):105-116.
11. Canning Division of General Practice Website: **Canning Data Extraction Tools**. Canning Division of General Practice Ltd; 2008 [http://www.canningdivision.com.au/dataextraction.html].
12. Australian Government, Department of Health and Ageing: **Series of National Guidelines (SoNGs) Influenza Infection (2010 Influenza Season) Guidelines for Australian Public Health Units**. [http://www.health.gov.au/internet/main/publishing.nsf/Content/cdna-song-influenza.htm].
13. Northern Sydney Central Coast Area Health Service: **Health-E-Profile**. 2010 [http://www.nscchhs.health.nsw.gov.au/publichealth/toc/choindex.htm].
14. Primary Health Care Research & Information Service: **GP numbers in New South Wales, 1999 to 2008**. [http://www.phcris.org.au/index.php].
15. Centers for Disease Control and Prevention: **Overview of influenza surveillance in the United States, Bethesda, Maryland**. 2010 [http://www.cdc.gov/flu/weekly/overview.htm].
16. Hope K, Merritt T, Eastwood K, Main K, Durrheim DN, Muscatello D, Todd K, Zheng W: **The public health value of emergency department syndromic surveillance following a natural disaster**. *Commun Dis Intell* 2008, **32**:92-93 [http://www.health.gov.au/internet/main/publishing.nsf/content/cda-cdi3201m.htm].
17. Nori A, Williams MA: **Pandemic preparedness. Risk management and infection control for all respiratory infection outbreaks**. *Aust Fam Physician* 2009, **38**(11):891-895[http://www.racgp.org.au/afp/200911/35810].

18. Zheng W, Aitken R, Muscatello M, Churches T: **Potential for early warning of viral influenza activity in the community by monitoring clinical diagnoses of influenza in hospital emergency departments.** *BMC Public Health* 2007, **7**:250[<http://www.biomedcentral.com/1471-2458/7/250>].

Pre-publication history

The pre-publication history for this paper can be accessed here:
<http://www.biomedcentral.com/1471-2458/11/435/prepub>

doi:10.1186/1471-2458-11-435

Cite this article as: Liljeqvist *et al.*: Automated data extraction from general practice records in an Australian setting: Trends in influenza-like illness in sentinel general practices and emergency departments. *BMC Public Health* 2011 **11**:435.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

