Flutracking surveillance: Comparing 2007 New South Wales results with laboratory confirmed influenza notifications

Sandra J Carlson, Craig B Dalton, Frank A Tuyl, David N Durrheim, John Fejsa, David J Muscatello, J Lynn Francis, Edouard Tursan d'Espaignet

Abstract

General practice and hospital surveillance for influenza-like illness (ILI) and laboratory influenza surveillance provide useful but incomplete information on influenza incidence. Flutracking is an Australian pilot of an Internet-based community ILI syndromic surveillance system designed to detect inter-pandemic and, potentially, pandemic influenza. Presence of fever and/or cough and absence from normal duties are collected weekly. Influenza vaccination status of respondents is recorded. New South Wales Flutracking data for 2007 were compared with New South Wales laboratory notifications for confirmed influenza to validate it's ability to provide alerts of influenza activity. Symptom rates amongst vaccinated and unvaccinated Flutracking respondents were compared using a variety of case definitions, with New South Wales laboratory influenza notifications. Time series methods were used to estimate the degree of correlation between each Flutracking case definition and the laboratory data. For the unvaccinated group, the correlations between all Flutracking case definitions and laboratory data were statistically significant, while for the vaccinated group no case definitions were significantly correlated with laboratory data. Thus Flutracking ILI data amongst unvaccinated participants correlated well with influenza laboratory surveillance. *Commun Dis Intell* 2009;33:323–326.

Keywords: influenza, surveillance, Flutracking, time series, ARIMA

Introduction

Seasonal influenza causes substantial morbidity and mortality each year.[1] Community-based surveillance of influenza-like illness (ILI) is therefore recommended by the World Health Organization (WHO) as part of a comprehensive influenza surveillance system during inter-pandemic and pandemic periods.[2,3] Influenza surveillance supports the detection and public health response to influenza transmission.[4]

It is acknowledged that while laboratory confirmed influenza surveillance data may be biased by testing activity, it is usually considered the most reliable indicator of the onset and peak of influenza activity. Therefore, laboratory data are often used as the default measure for comparing the performance of syndromal (or 'syndromic') influenza surveillance. Zheng et al compared emergency department visits assigned a clinical diagnosis of influenza to New South Wales influenza laboratory data to determine whether the former could offer earlier warning of an increase in influenza incidence in the New South Wales population.[5] Lau et al defined the start of peak influenza activity using laboratory isolation rates for their analysis of multiple streams of influenza surveillance data.[6]

Flutracking is a weekly community online survey of ILI that integrates syndromic information with participants' influenza immunity status. Flutracking aims to help fill the gap between laboratory and syndromal surveillance systems because it uniquely combines information on influenza symptom rates and vaccination status of participants. It has been piloted with approximately 900 participants predominantly in New South Wales in 2007 and this rose to over 4,000 nationwide in 2008.

The purpose of this study was to use sound time series methods to validate the 2007 New South Wales Flutracking data against New South Wales data for laboratory confirmed influenza.

Methods

Flutracking recruitment

Flutracking was initially piloted in 2006. Recruitment occurred as outlined in Dalton et al.[7] Potential participants were directed to a web page providing information about the study and an online consent form. A confirmatory email response from the participant's email address was required to complete enrolment. The study was approved by the Hunter New England Area Health Service Human Research Ethics Committee. Participants were allowed to join at any time during the surveillance period.

Flutracking data collection

Each Monday from 4 June to 15 October 2007, participants received an automatically generated weekly email link to the online questionnaire. In the 1st online questionnaire participants were asked

about their usual postcode of residence; whether they work face-to-face with patients in hospitals, nursing homes, doctors' surgeries or as community health workers; their month and year of birth; and whether they received an influenza vaccination in the previous or current year.

For each subsequent questionnaire, participants were asked whether during the prior week (ending Sunday) they had experienced fever and/or cough and/or muscle aches on any specific day/s, and whether they had been absent from usual activities on any specific day/s. Participants who reported not being vaccinated against influenza in the current season were asked if they had received vaccination in the prior week during each weekly survey. If they responded in the affirmative the question was automatically deleted from their subsequent weekly surveys.

Analysis

Data for New South Wales participants for the week ending 3 June 2007 to the week ending 14 October 2007 were included in the analysis. New South Wales data accounted for 76% of all participants in Australia who completed at least 1 survey during 2007. For the purpose of this analysis, the laboratory data was classified as the independent variable, and each of the Flutracking symptoms were classified as dependent variables.

For each of the vaccinated and unvaccinated groups, a time series of the proportion of respondents reporting any of 5 possible case definitions was created. The case definitions were:

- fever only;
- cough only;
- absence from work or normal duties;
- fever and cough; or
- fever, cough and absence from work or normal duties.

A time series of weekly counts of positive influenza antigen tests (polymerase chain reaction and direct immunofluorescence) were created from the NSW Department of Health notifiable diseases database.[8] Counts were aggregated into weeks based on the date of specimen collection.

We used autoregressive integrated moving average (ARIMA) time series and cross correlation analysis to determine whether there was an association between the laboratory time series and weekly proportions for each Flutracking case definition. As the Flutracking data used for analysis were proportions the variance stabilising transformation for binomial data was applied.[9] This is an arcsine transformation, $\gamma_a = \arcsin\sqrt{\gamma}$, where γ_a is the transformed Flutracking data, and γ is the proportion of participants with the particular Flutracking symptom/s specified by each case definition. Similarly, the laboratory data were counts, and the variance stabilising transformation for a Poisson distribution was applied: $x_a = \sqrt{x}$, where x is the original laboratory data, and x_a is the transformed laboratory data.[9]

In time series modelling, the assumption that model residuals are independent is typically violated due to the residuals being autocorrelated (i.e. the current values of a series correlate with past values of the same series).[10] If autocorrelation is not removed, then the relationship between 2 time series could be overestimated.[11] Any comparisons made between laboratory data and Flutracking data potentially require correction for autocorrelation.

For the vaccinated and unvaccinated groups, we calculated raw correlations and used ARIMA models to estimate the association between weekly proportions of respondents reporting each case definition and weekly counts of positive influenza isolates. ARIMA modelling is a well established time series analysis technique that can be used to model an autocorrelated variable.[10] Adding an independent variable to the usual ARIMA model (called transfer function analysis)[12] allows the relationship between 2 time series to be measured, while correcting for autocorrelation. The SAS ARIMA[13] procedure was used to compute cross correlations between the 2 data series at various time differences, after both series had been 'prewhitened' (that is, filtered by an ARIMA model that was originally fitted to the independent variable).

Results

Descriptive statistics

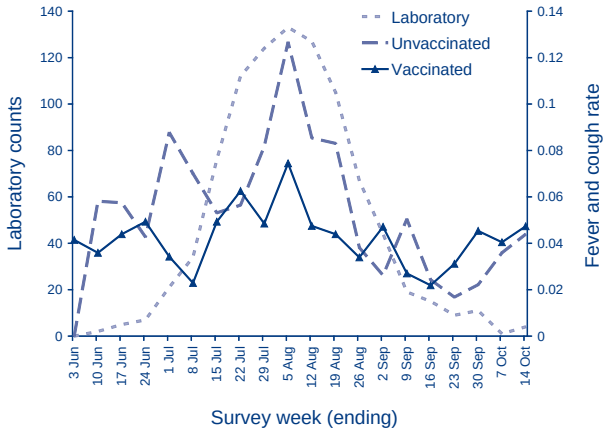
In New South Wales, for the 20 week period between 3 June and 14 October 2007, there was an average of 502 participants per week who completed the survey. Over that period, a weekly average of 65% of participants reported being vaccinated.

Visual inspection of the time series of each Flutracking case definition against laboratory data suggested that the peaks in laboratory data corresponded to periods of high Flutracking symptom rates for the unvaccinated group compared with the vaccinated group. A graph for the 'fever and cough' case definition is shown in the Figure.

Raw correlation analysis

Using raw correlation analysis (i.e. without autocorrelation correction), we found that the correlation values were generally highest when Flutracking

Figure: Flutracking symptom rates for 'fever and cough' case definition, compared with influenza laboratory notification counts, New South Wales, 2007, by influenza vaccination status



symptom rates and laboratory data were compared in the same week (i.e. a lag of 0), but similar values also occurred at other differences in time (or lags).

Each Flutracking case definition in both the vaccinated and unvaccinated groups showed a statistically significant relationship with the laboratory data at a lag of zero (all *P* values for the correlation coefficients were less than 0.05). However, it was important to further analyse the relationship between the two time series using ARIMA analysis.

Autoregressive integrated moving average analysis

Results from an autocorrelation check for white noise using ARIMA analysis indicated that laboratory data showed significant autocorrelation (at the level of $P = 0.05$), and that the model that fitted this data best was $\gamma_t = 1.6\,\gamma_{t-1} - 0.6\,\gamma_{t-2} + \varepsilon_t$ where γ_t is the laboratory data at time t (in weeks), and ε are the residuals from the model. This model was used to pre-whiten both the Flutracking and laboratory data.

Cross correlations for the residuals from the ARIMA model applied to the laboratory data and each of the Flutracking data series are summarised in the Table. Only cross correlation values at a lag of zero for each case definition related to laboratory data are reported.

In the unvaccinated group, all cross correlations at a lag of 0 weeks were statistically significant at a level of $P = 0.05$. The cross correlation analysis did not provide evidence of a substantive difference between the case definitions, except for 'absence from work or normal activities,' which at 0.442, did not have as high a cross correlation as the other symptoms. In the vaccinated group no case definitions at a lag of zero were statistically significant at a level of $P = 0.05$. The results from the ARIMA analysis for the vaccinated group were not consistent with results from raw correlation analysis, where there were statistically significant relationships between every case definition for the vaccinated group and the laboratory data.

Discussion

There was a statistically significant correlation between time series of laboratory confirmed influenza and Flutracking data for unvaccinated participants in New South Wales for all 5 case definitions (fever; cough; absence; fever and cough; fever, cough and absence) at a lag of 0 weeks. This indicates that Flutracking responds contemporaneously with laboratory surveillance of disease caused by influenza that leads to a specimen being col-

Table: Cross correlation and corresponding probability values from the ARIMA analysis for each Flutracking case definition symptom rate compared with influenza laboratory notifications, New South Wales, 2007, by vaccination status

Vaccination status	Case definition	Cross correlation value	Probability value for cross correlation (using a one-tailed t statistic)
Vaccinated	Fever	–0.006	1
Vaccinated	Cough	0.302	0.097
Vaccinated	Absence	–0.054	1
Vaccinated	Fever and cough	0.203	0.188
Vaccinated	Fever, cough and absence	–0.072	1
Unvaccinated	Fever	0.654	0.005
Unvaccinated	Cough	0.623	0.006
Unvaccinated	Absence	0.442	0.032
Unvaccinated	Fever and cough	0.640	0.005
Unvaccinated	Fever, cough and absence	0.652	0.005

lected. For the vaccinated group who should have at least some protection against influenza infection, cross correlations were not statistically significant after correction for autocorrelation, indicating that Flutracking can discriminate between influenza and other causes of ILI disease.

For vaccinated participants, the change in statistical significance between raw correlation results and ARIMA modelling results demonstrates the importance of adjusting for autocorrelation, and using appropriate analysis techniques for time series data. Without controlling for autocorrelation, spurious results were obtained. However, after correcting for autocorrelation the 'true' relationship between the 2 data series could be seen.

A limitation when quantifying the relationship between the Flutracking and laboratory data was that there were only 20 continuous time points in the weekly Flutracking data series, when usually at least double that number are recommended for ARIMA analysis.[10] However, we confirmed by Monte Carlo simulation that a model of the type found for the laboratory data, nearly always generates data that are clearly autocorrelated, even when there are only 20 time points, based on checking by time series analysis.

In conclusion, this analysis of Flutracking results has provided support for its value in providing alerts of influenza activity. Distinguishing between vaccinated and unvaccinated participants offers further potential to determine the value of Flutracking in assessing the effectiveness of the annual influenza vaccine composition in real-time.

Acknowledgements

Sandra Carlson drafted the manuscript and performed the statistical analysis. Craig Dalton conceived and designed the Flutracking project, contributed to the statistical analysis and contributed to the manuscript. Frank Tuyl oversaw the statistical analysis and contributed to the writing of the manuscript. David Durrheim contributed to the design of the Flutracking project and writing of the manuscript. John Fejsa contributed to the design of the Flutracking project and had primary responsibility for the online software and database development, as well as questionnaire design. David Muscatello contributed to the statistical analysis and writing of the manuscript. Lynn Francis contributed to the statistical analysis and writing of the manuscript. Edouard Tursan d'Espaignet contributed to the design of the Flutracking project, statistical analysis, and writing of the manuscript.

The authors would like to thank Robin Gilmour for providing New South Wales influenza laboratory data, and to acknowledge the University of Newcastle, NSW Department of Health, and the Hunter Medical Research Institute for their continued support. We would also like to acknowledge the thousands of Flutracking participants who give their time freely each week to contribute to influenza surveillance.

Author details

Ms Sandra J Carlson, Trainee Biostatistician[1,2]
Dr Craig B Dalton, Public Health Physician[1]
Dr Frank A Tuyl, Statistician[1]
Professor David N Durrheim, Health Protection Service Director[1]
Mr John Fejsa, Applications Developer[1]
Mr David J Muscatello, Senior Epidemiologist[2]
Dr J Lynn Francis, Statistician[1]
Dr Edouard Tursan d'Espaignet, Surveillance and Monitoring Service Director[1]

1. Hunter New England Population Health, Wallsend, New South Wales
2. NSW Department of Health, North Sydney, New South Wales

Corresponding author: Ms Sandra Carlson, Trainee Biostatistician, Hunter New England Population Health, Locked Bag 10, WALLSEND NSW 2287. Telephone: +61 2 4924 6406. Facsimile: +61 2 4924 6029. Email: sandra.carlson@hnehealth.nsw.gov.au

References

1. Pitman RJ, Melegaro A, Gelb D, Siddiqui MR, Gay NJ, Edmunds WJ. Assessing the burden of influenza and other respiratory infections in England and Wales. *J Infect* 2007;54:530–538.
2. World Health Organization. WHO global influenza preparedness plan. The role of WHO and recommendations for national measures before and during pandemics. WHO/CDS/CSR/GIP/2005.5. Geneva: World Health Organization; 2005. Accessed on 21 October 2008. Available from: http://www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP_2005_5/en/index.html
3. Teutsch SM, Churchill RE, eds. *Principles and Practice of Public Health Surveillance*. 2nd edn. Oxford, New York: Oxford University Press; 2000.
4. O'Brien K, Barr IG. Annual report of the National Influenza Surveillance Scheme, 2006. *Commun Dis Intell* 2007;31:167–179.
5. Zheng W, Aitken R, Muscatello DJ, 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.
6. Lau EH, Cowling BJ, Ho LM, Leung GM. Optimizing use of multistream influenza sentinel surveillance data. *Emerg Infect Dis* 2008 Jul [Online] Accessed 21 October 2008. Available from: http://www.cdc.gov/EID/content/14/7/1154.htm
7. Dalton C, Durrheim D, Fejsa J, Francis L, Carlson S, Tursan d'Espaignet E, et al. Flutracking: A weekly Australian community online survey of influenza-like illness in 2006, 2007 and 2008. *Commun Dis Intell* 2009;33(3):316–322.
8. NSW Department of Health. Notifiable Diseases Database (HOIST), Centre for Epidemiology and Research.

9. Fisher LD, van Belle G. Biostatistics: A methodology for the health sciences. New York: John Wiley & Sons Inc; 1993.
10. Pankratz A. Forecasting with dynamic regression models. New York: John Wiley & Sons Inc; 1991.
11. Schwartz J, Spix C, Touloumi G, Bachárová L, Barumamdzadeh T, le Tertre A, et al. Methodological issues in studies of air pollution and daily counts of death or hospital admissions. *J Epidemiol Community Health* 1996;50 (Suppl 1): S3–S11.
12. Box GEP, Jenkins GM, Reinsal GC. Time series analysis: Forecasting and control. New Jersey: Prentice Hall; 1994.
13. SAS 9.1.3 Help and Documentation. Cary (USA): SAS Institute; 2002–2004.