Dear Sir/Madam

Potential for early warning of viral influenza activity in the community by monitoring clinical diagnoses of influenza in hospital emergency departments

I am writing to provide a revised manuscript in response to reviewer comments. We are grateful to the reviewers for providing thorough and valuable feedback.

Reviewer comments are shown in Blue. Our responses are shown in black following the reviewers' comments.

**REVIEWER 1**

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1) What is the true value of a 3 day lead time in surveillance? The authors need to justify the public health relevance of such a lead time. During normal epidemic periods, how would interventions change with this slightly earlier warning? In a pandemic crisis, what value would three days provide? I think there may be some clear benefits but the authors should be more explicit here. While the authors discuss the additional reporting delays, they also acknowledge that these will be reduced through changes in reporting structure. The public health importance of the findings needs to be clearly addressed in the discussion.

The public health benefit of the early warning provided by ED data is to provide “situational awareness” to public health professionals at least three days earlier than would otherwise be available. It is unlikely that public health professionals would treat influenza virus as the certain cause of a syndromic surveillance signal. However, the time advantage is particularly important in rejecting the worst scenarios of bioterrorism with an agent that mimics influenza in its prodromal stage, possible circulation of a pandemic strain or seasonal vaccine failure. This would be facilitated, for example, by triggering a cascade of events aimed at identifying the causative organism, such as increasing the number of tests to be ordered for patients with ILI, possibly applying a wider range of tests than would otherwise be performed, and then using that information to guide subsequent public health prevention and control efforts.

In relation to influenza generally, the propensity for the virus to capriciously mutate combined with its short incubation period means that any early warning is advantageous. It is difficult to prospectively know whether a mutated strain (pandemic or otherwise) has commenced circulating and that the population is well covered by the existing vaccine schedule. This is particularly important outside of the
influenza season when there is a low index of suspicion for influenza but when outbreak have occurred.

Regarding pandemic influenza, in many countries a pandemic strain is likely to be detected independently of ED surveillance through epidemiological suspicion in travelers, but detection prior to circulation is not guaranteed. Therefore, targeted testing signaled by syndromic surveillance can provide an advantage for pandemic preparedness. Further, pandemics often occur in waves, and should a pandemic commence, early warning of a second wave would be an advantage, particularly as laboratory testing may have declined once the pandemic is well established. We therefore believe the early warning benefit is valuable in a wide range of public health situations.

We have extended some discussions on the potential public health benefit as part of the second paragraph on page 14.

2) One important oversight is that the authors claim originality in their analysis of ED data against a continuously collected laboratory standard. Consequently, there have been several studies that have used lab-based data for evaluation of syndromic surveillance. These include:

   These references should be cited and the authors should clearly define what advance this analysis provides (i.e. analysis of temporal relationships, timeliness.). The use of daily analysis is fairly novel yet the public health value of daily measures is not clear unless in a pandemic period (see above). Clearly, additional analysis will strengthen support for syndromic surveillance of influenza-like illness but the authors should provide better justification of the paper’s scientific contribution.

   In the introduction we have referenced these papers and more clearly elicited the rationale for our study (added the second paragraph on page 4).

   Many syndromic surveillance systems analyse data on a daily basis, thereby giving our analysis direct relevance (added as part of the first paragraph on page 16).

   Please refer to the previous point regarding the public health value.

3) How have the authors dealt with the co-circulation of other respiratory viruses (including RSV)? Because other respiratory illness epidemics may peak earlier than influenza epidemics this may bias the analysis toward earlier signaling of ED data.

   Our study was aimed at assessing co-movement of the incidence of influenza clinical diagnoses and laboratory results for influenza virus. It is possible that RSV and other respiratory viruses caused some of the movement in the ED time series, however, it would be unlikely that such a consistent correlation in the most recent three years of our study could be caused by other organisms when we're comparing with an influenza laboratory standard.
Also, Bourgeois et al 2006 found that the timing and shape of an RSV time series differed from that of an influenza series. For the majority of the years studied, the change in incidence of RSV positivity rates occurred at markedly different times from changes in incidence of influenza positivity rates.

In our study, the broad shape of the ED data was similar to that of the laboratory data. However, we do acknowledge that the ED data may be contaminated by other respiratory viruses.

We have added additional discussion as part of the first paragraph on page 18.

4) The statistical framework for evaluating ED data is novel and interesting. However, validation of this method would be nice. For instance, the authors should consider looking at peaking of the raw time series data as an additional measure of epidemic timing. There is a relatively simple yet reliable validation approach.

To demonstrate the plausibility that the timeliness of ED series relative to the laboratory series is real, we have added an additional figure (Figure 5). The 7-day moving average counts demonstrate the clear lag between the rise, peak and decline of the two data sources.

We have edited the methods (the third paragraph on page 9), results (the second paragraph on page 12) and discussion (the first paragraph on page 13).

5) The correlations reported by the authors do not appear very convincing. The authors state that these are statistically significant positive correlations but P-values should be provided.

We have added a column showing P-values in table 3 and revised the relevant text in the results (the second paragraph on page 11). We believe the consistency of the three-day correlation over the most recent three years of data in our study provides further support for a real effect.

The correlation is based on a day-for-day match between the two time series. However, the actual lag is likely to be variable depending on the stage of illness that people sought medical attention. The cross-correlation graph we have now included (Figure 3), shows a cluster of positive correlations around three days, and this probably is an artifact of this variable delay (added as part of the second paragraph on page 13).

-----------------------------------------------------------------------------------------------
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
-----------------------------------------------------------------------------------------------
Discretionary Revisions (which the author can choose to ignore)
1) How did the authors select ICD9 code 487 for classification? There is a broader range of codes have been used in the past and may be appropriate (See Mardsen-Haug, 2007)
We have used these ICD9 and ICD10 codes over a number of years and have found them to be most useful for monitoring influenza like-illness related to influenza circulation. This is because the influenza syndrome has a broadly similar pattern to the laboratory positive results for influenza in terms of the timing, shape and magnitude of peak influenza activity compared with other respiratory syndromes that we have tried. This study provides some validation of this decision. We have added justification for our choice in the methods section on page 6 and added an additional reference (No. 30).

The findings of the very recent study by Marsden-Haug may be useful for future surveillance evaluation studies.

2) The authors should provide more detail about the source of the laboratory data. Is this the same source of patients as the ED data?

Laboratory flu notifications are from all public or private laboratories across the state. The majority of the laboratory notifications were from private laboratories and this means that these tests were ordered for patients presenting to GP clinics, private hospitals and other non-public health care services, such as nursing homes. The ED data ONLY include ED visits presenting to EDs of the public hospitals. Only 5% of ED visits were admitted to an inpatient ward. Tests for flu are unlikely to be ordered for those who have uncomplicated clinical signs and are admitted to ward. Therefore, the ED and laboratory data do not largely share the same source of the patients but we do acknowledge that there may be some overlap.

We have revised the methods (under the heading “Laboratory data” on page 7) and the discussion (the first paragraph on page 14) to clarify that the ED and laboratory data do not largely share the same source of patients.

3) We appreciate the author’s citation of our previous work in ED surveillance of influenza (AJE 2005). We would just like to point out while the Serfling regression is a method for estimating relative impact of influenza epidemics (not timing) by removing the seasonality during non-influenza periods, the Fourier transformation applied in the AJE paper explicitly incorporated the influenza epidemics to estimate timing. The authors should take a look at a number of recent studies analyzing mortality data using similar spectral decomposition methods based on the annual series (e.g.: Viboud et al. 2006. Science) for defining the timing of influenza epidemics.

We acknowledge the sensitivity of the reviewer's position in assessing our paper. However, we feel that the point is an important one and we believe it to be correct. The excess mortality above the seasonal influenza and pneumonia mortality background is widely used to estimate influenza-associated mortality during the flu season. As Viboud et al's paper (2006) pointed out, this excess mortality is a useful indicator of the timing of the epidemics. Therefore, we believe when assessing the timeliness from pneumonia and influenza mortality trends, it is more important to compare the short-term excess incidence over and above the seasonal background rather than the seasonal background itself (eg. yearly cycle) which could be driven by
many different organisms and factors other than influenza. The paper by Bourgeois et al (2006) mentioned earlier also demonstrates that RSV and many other organisms are contributors to seasonal respiratory activity. We believe that the background seasonal component of pneumonia and influenza mortality trends is not a reliable indicator of influenza activity.

We have referenced the paper by Viboud et al and revised the discussion on page 16 to make our argument clearer.

**REVIEWER 2**

**Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)**

These compulsory revisions need not require a lot of time, but:

1) Is there a practical suggestion to detect the onset of influenza season using ED data and to show that ED data would typically detect the season before the positive test results tend to? (note that I think the number of test orders and/or the fraction of positive results might complement the no. of positive results)

My understanding: at the end of each year you fit a spline to capture longer-term trends (seasonality mainly) and other terms to capture daily effects retrospectively to the entire year. This is useful to determine, as you did, that there might be a tendency to see large positive forecast errors in the ED data prior to the test orders (but see my next comment). This approach does not lend itself to actually detecting the onset of flu season more quickly (and how might the onset be defined?)

The method we used in this study is not intended to prospectively detect the start of an epidemic but is designed to retrospectively measure the relative timeliness between two time series. To detect influenza outbreaks (seasonal or unseasonal) from the time series, outbreak detection techniques, such as cumulative sum or exponentially weighted moving averages, are required (eg. Sonesson and Bock, 2003). To make the aims of the paper clearer, we’ve altered the manuscript title, abstract, last paragraph of introduction (page 5), discussion (first paragraph on page 13), and conclusions.

We appreciate the potential value of monitoring the number of test orders, but it is outside the scope of this study because year-round information on test orders is not readily available in our state. We have added a reference to this possibility in the Discussion (at the end of the second paragraph on page 15).

2) The ED residuals cannot truly be white noise (in practice, they almost never are) or there would be no cross correlation with the positive lab results residuals. This is OK, but make this point clearly in the revision, and: what lags were examined and do small but statistically significant correlations at negative lags seem to arise due to actions happening near the onset of the flu season(s)?

Please provide statements regarding exploratory evaluation to convince me that this very small but statistically significant (unless a lot of lags were examined and you didn't property account for multiple testing??) correlation is "real" and explainable (arising, say due to "action" associated with the beginning of the annual peak) and
therefore potentially exploitable (how might it be exploited in "real time", per my remarks in (1) above). If you convince me that this is "real," I'd settle for remarks involving "we have no practical suggestion" for how to actually exploit this tendency, but the retrospective approach (involving retrospective spline fitting to capture the long term trend) reveals a real pattern. Basically: I'm not yet convinced that you found a real pattern and I don't see why a lag of 3 days would occur if you date lab results to the day of lab order.

We have altered the methods section to avoid giving the false impression that the residual time series were completely white noise (the second paragraph on page 8)

Up to 30 lags in both directions were examined and we have revised the manuscript (first paragraph on page 9) and added an additional figure (Figure 3) showing the cross-correlation plot of the residuals from the models for the entire five-year period.

We think that the small and statistically significant correlation does represent a real phenomenon that the short-term changes in ED series tended to be followed by the short-term changes in laboratory series. The consistency in the cross-correlation results across each of the most recent years (3-4 days in 2003,2004,2005) supports that this finding did not occur by chance (Table 3) (added as part of the last paragraph on page 12). We believe the most likely reason for the difference in movement of the two time series would be that respiratory specimens are more likely to be taken from patients with more advanced or more serious illness or those who are experiencing secondary complications of infection such as pneumonia. The illness may take longer to progress to a stage where a clinician decides that a laboratory test is useful in clinical management. We believe that otherwise healthy patients with a classic influenza-like syndrome are more likely to be assigned the provisional influenza diagnosis and are also unlikely to have a specimen taken. This offers another explanation for the small correlation, because the duration of infection prior to a specimen being taken would vary substantially and thus we would not expect the strongest correlation to fall at one exact lag only. In fact, the cross-correlation graph shows a cluster of small positive correlations around lag 3 days that could reflect this variability (added as part of the second paragraph on page 13).

Other possible reasons for the small correlation include the difference in scope between the ED and laboratory data sources: laboratory results are drawn from both public and private laboratories for GP clinics, patients of public hospitals (ED and admitted), private hospitals (EDs are rare in private hospitals in Australia) and other non-public health care services, such as nursing homes. ED visits are drawn only from public hospital EDs. In fact, the majority of influenza results come from private laboratories that do not service public EDs. This difference is reflected in the different age structure and geographic distribution shown in Table 1 (added as part of the first paragraph on page 14).

To demonstrate the plausibility that the timeliness of ED series relative to the laboratory series is real, we have added an additional figure (Figure 5). The 7-day moving average counts demonstrate the clear lag in the rise, peak and decline of the two data sources. We have edited the methods (the third paragraph on page 9), results (the second paragraph on page 12) and discussion (the first paragraph on page 13).
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Fig 2 vs Fig 3:
I cannot see how the log of the spline fit should be so large (approx 2 in some instances) when the maximum count is approx 60. Do the other terms in the "quasi poisson" regression add negative contributions?

The negative contributions in the regression came largely from a strong negative day-of-week effect in the ED series and a negative intercept in the laboratory series.

The day of week variable is a categorical variable range from 1 (Sunday) to 7 (Saturday). The reference date for the day of week variable is 1 (Sunday), which is one of the busiest days for EDs. As a result, the estimates for the rest of weekdays (2-7) were negative in ED models.

We would prefer not to alter the text of the manuscript regarding this point because it may confuse some readers and would not change our conclusions.

Also, write the "final model" as Expected(log(counts)) or instead, add an error term.

Thank you - we’ve revised the formula on page 8.

Any guess why you can't remove temporal correlation from all series?

There are residual autocorrelations in the ED residual series in 2002 and the laboratory residual series in 2005.

Essentially we believe the autocorrelation in infectious disease time series derives largely from the communicability of the organism, such that each person that becomes infected has the opportunity to infect others and so on. This cause successive incidence observations to be serially correlated.

In 2002, there was a prolonged but not sharp flu season. This extended period of infectivity may have influenced the overall autocorrelation for the year. We are unable to say why stronger autocorrelation wasn't apparent in the laboratory data for that year, other than the possibility of incomplete reporting occurring in the early years of the requirement for laboratories to notify influenza.

In 2005, there was a known data problem with the laboratory data where some false positive test results were reported from a single laboratory. We think this may have influenced the autocorrelation in that year.

We could have controlled autocorrelation in those years by increasing the number of knots for the spline from 11/year to 14/year. However, we made a prior decision to choose the minimum degree of smoothing that would adequately remove autocorrelation from the two five-year time series. This facilitated consistent interpretation of our results and avoided overfitting the models.
We have added the above points in the Discussion on the first paragraph on page 16.

Discretionary Revisions (which the author can choose to ignore)

Explain (conjecture?) why you seem to find that the positive influenza test results lag the ED influenza by 3-4 days even though you use the date of test orders rather than the date of the positive test.

This was addressed in the 2nd point of the Major Compulsory revisions above.

Can you say anything about the tendencies for more or less test orders as a function of time or year? I assume there is no "sentinel physician" concept in which influenza tests are routinely ordered throughout the year simply for flu surveillance rather than because of the MD's best guess that influenza might be the illness.

Related: might there be a tendency for more test orders after influenza seems to be on the rise?
Related: why not record the number of lab tests, and the number of positive results. You are using only the number of positive results, but also available in perhaps more timely manner is the number of test orders.

We cannot discount the possibility that more tests are ordered when knowledge that flu is circulating becomes commonly known. However, the laboratory notification system is a passive surveillance system and active sampling and sentinel surveillance has declined in our state over recent years. On the other hand if clinicians know that flu is circulating they may not choose to test because influenza has a higher probability of being the cause of illness. Also, during off-season, adults with ILI may be more likely to be tested than during the flu season because influenza as a cause is unexpected. We have expanded the discussion of this limitation in the second paragraph on page 19.

It is a good idea to monitor the number of test orders as it may be used as a proxy for the level of flu activity in the population. However, as stated earlier, information on test orders is not routinely captured in any year-around surveillance system in our state. We also do not want to further extend the scope of the present study.

I hope these changes are satisfactory and look forward to your response.

Yours sincerely

Wei Zheng
Senior Surveillance Officer
Centre for Epidemiology and Research
6 June 2007