Author's response to reviews

Title: Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey

Authors:

Dena L Schanzer (dena.schanzer@phac-aspc.gc.ca)
Hui Zheng (hui.zheng@phac-aspc.gc.ca)
Jason Gilmore (Jason.Gilmore@statcan.gc.ca)

Version: 2 **Date:** 20 December 2010

Author's response to reviews: see over

Ref: MS: 1233077791451727
Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey
Dena L Schanzer, Hui Zheng and Jason Gilmore

20 December, 2010

Dear Dr Marianne A.B. van der Sande, Ms Roxane Rajabi
Editors, BioMed Central

On behalf of my co-authors, I would like to thank you for your expressed interest in our manuscript and especially the reviewers for their helpful comments and suggestions concerning our manuscript. Revisions have been made to the manuscript as per the comments of the reviewers and highlighted using the track changes option. The responses to reviewer comments have also been outlined below.

Yours sincerely,

Dena Schanzer
Corresponding author
Public Health Agency of Canada

Editorial Requests:

Please can you clarify whether any permissions or ethical approval was needed for the use of the data in your study?

Ethnical approval was not required for the data used in this study. The Labour Force Data came from Statistics Canada under a cost recovery agreement. Ethical approval is not part of the STC protocols. Permission to use the data provided by FluWatch was obtained from the FluWatch team. Representatives from both organizations are on the author list.

Reviewer's report
Title: Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey
Version: 1 **Date:** 6 November 2010
Reviewer: Jeff C Kwong
Reviewer's report:
This is an interesting and original paper that used data from Canada's Labour Force Survey and laboratory surveillance data to estimate work absenteeism attributable to seasonal and pandemic influenza through statistical models.

Major Compulsory Revisions:

Methods
1. *Additional description of the Labour Force Survey would be helpful (e.g., sampling strategy, sample size, etc)*

Author response:

Additional information about the LFS has been included in the revised version.

2. *The model includes a term (FluAr) for the number of weekly influenza A confirmations and another (fluBpp) for the percent of tests positive for influenza B. The authors should provide justification for using the number of confirmations for one influenza type and the percent of tests positive for the other – why the inconsistency?*

Author response:

The short answer is that the effect of influenza B on absenteeism could not be estimated separately for each season due to limited statistical power and/or impact of influenza B on absenteeism, so that the use of *fluBpp* was one approach to account for the effects of influenza B in the model. The results of various model forms were reviewed to see if other respiratory viruses (RVS, adenovirus or parainfluenza) might have a significant impact on absenteeism/ hours lost. Only the results for influenza B were statistically significant, and hence included in the final model described in the paper.

Use of the number of influenza positive A tests along with separate parameters for each season (rather than the percent positive), in addition to providing a slightly better model fit, provides a redundancy that helps illustrate model robustness. While there were statistically significant differences between the beta3s by season, the beta3s were similar enough that the burden attributable to influenza for each season was plausible. Overfitting becomes a problem when parameter estimates start to vary widely along with a noted loss statistical significance. The additional parameters, one for each season, allowed for independent estimates of the

influenza attributed disease burden for each season. The results are presented in Figure 1 and help support the validity of the results.

There are various models that have been used to estimate influenza-attributable events or excess mortality. All regression models include variables to explain the weekly or monthly seasonality and secular trends of the dependent variable and to account for the impact of influenza activity. Various proxy variables for influenza have been used including the number of influenza deaths, the proportion of tests positive for influenza, ILI consultation rates, and the number of influenza positive tests. All are strongly correlated, though with subtle differences. "Percent positive" is a simple approach to normalizing the number of positive tests for differences in background testing rates. As with any proportion, *fluBpp* can be influenced by a number of external factors, such as the circulation of other respiratory virus that cause ILI.

Most published models have used the simpler parameterization, sinusoidal seasonality; 1 trend variable; and *fluApp*, *FluBpp*, with perhaps *RSVpp*. Generally the information content of the dependant variable and proxy variables for influenza activity have been sufficient to support the fuller parameterization used in this model. Where over fitting becomes apparent is when parameter estimates for other respiratory viruses lose face validity (large variation in parameter estimates for one season compared to another, very large confidence intervals, estimates are positive for one season and negative for another). As well the fuller model shows that the seasonally of the dependent variable is not necessarily sinusoidal, secular trends in the dependant variable are not necessarily linear, and the relationship of the proxy variable for influenza activity and the dependant variable varies yearly (by flu season). A degree of consistency in the parameter estimates from season to season provides face validity of model results and was one criteria guiding model development.

3. Why is #5 only for the pandemic year – why is there not such a term for other seasons? I'm also not sure it should be called "non-influenza related ILI" – "non-influenza related illness" would probably be more appropriate since we don't know that the work absenteeism or hours lost are necessarily associated with ILI (i.e., fever, cough, etc.)

Author response:

The variable *Pandemic2009 (Beta#5)* was included in the model because there were a number of reasons to suspect a change in behaviour during the 2009 pandemic – that is changes in behaviour in comparison to the

typical reaction of seasonal influenza. Public health messages throughout the pandemic period encouraged all Canadians "to stay home if sick". In other work, there was some evidence that, for example, ILI related physician visits increased throughout the pandemic period, and that there was a component of the increase that was not related to the changing level of influenza activity. On the other hand, discretional sick leave may have been reduced, due to concern that extra sick days may be needed because of the pandemic. The actual behavioural change was likely more complex than accounted for by a single indicator variable.

The explanation has been reworded to stress that this parameter would account for behavioural change related to the public announcement of the pandemic rather than infection with the pandemic strain of influenza.

4. *What is the rationale for including a term (Hospadmsr) for hospitalizations?*

Author response:

The term *Hospadmsr* (Beta#6) (laboratory confirmed pH1N1 hospital admissions for the reference week) was included in the model because this series was considered the best proxy for the level of influenza activity during the pandemic. Laboratory testing practices changed significantly throughout the pandemic period and the number of confirmed cases was not considered a good proxy for the level of influenza activity. Initially, each community engaged in heightened testing in order to identify community transmission in the early stages of the pandemic. In following these early cases, public health quickly realized that this pandemic strain was relatively mild in most cases. Once community transmission was identified, intensive case identification was no longer a priority, and testing was limited primarily to hospitalized patients in many jurisdictions.

This explanation was added to the statistical analysis section.

Results

5. Results, first two sentences: it's not clear to me how these estimates were obtained. My understanding of the model is that it estimates the proportion of working individuals who experience absenteeism for the particular week of each month. But how are the final estimates contained in these two sentences generated? Is it by summing the absenteeism rates for each month over the course of each influenza season? If so, doesn't that assume that people can be absent from work only once (i.e., that the individuals absent each month are mutually exclusive)? What about people who have more than one absence during an influenza season? Further explanation of this and/or clearer language are needed.

Author response:

Yes, this figure was calculated by summing the predicted absenteeism rates for each season or indicated period. Dual influenza infections in one employee (and in different months) would be counted as two absences. The absenteeism rate was pro-rated to the full month based on the number of work days in the month with an adjustment for variation in the level of influenza throughout the month.

The denominator for absenteeism rates was the average number of employed persons for the specified period. The proportion of hours lost due to influenza was calculated by summing the estimated hours lost due to influenza and the potential hours worked over the specified period and then dividing.

A description of these calculations has been added to the methods section.

Discussion

6. First sentence: "estimates of absenteeism due to seasonal influenza ranged from 5% to 20%" – this is not presented anywhere in the results; it is not clear where these numbers came from.

Author response:

The specified range was from results presented in Figure 1 based on the 10 and 90 percentiles. The estimate for 2005/06 (A/California/7/2004) was not statistically significant, so it is difficult to say where the exact lower limit is. The inter-quartile ranges (Q1 - Q3) as well as the 10 and 90 percentiles have been added to the results section, and additional discussions of these limitations have been included in the discussion section.

7. Might some of the hours lost during the pandemic be attributable to time required to get vaccinated or to isolation (either imposed by the employer or self)? If so, there might be some benefits associated with the absenteeism (i.e. reduced transmission of influenza). The latter might be applicable as well to seasonal influenza. The authors did not seem to consider this.

Author response:

Hours lost for vaccination would be captured in the seasonal baseline, and/or contribute to the unexplained variation.

The effects of isolation (stay home if sick) have been discussed in more depth.

Benefits associated with absenteeism were not addressed in this study and should not affect the relationship between the weekly number of lab

confirmed cases and the weekly number of influenza cases/absences due to influenza infection.

8. Last paragraph, first sentence: it is unclear how the authors conclude that "an upward trend in absenteeism rates due to seasonal influenza was noted for recent years." Is that based on Figure 4? While it is fairly evident that absenteeism rates are increasing over time, it is not obvious that absenteeism rates attributable to influenza are increasing (gray line).

Author response:

We did not include a table or figure showing this upward trend by year/season. This is partly shown in Figure 1 as noted. (As noted, Figure 4 shows the upward trend in all absenteeism, not influenza-attributable absenteeism.) The four influenza seasons where a single antigenic strain dominated (with minimal co-circulation of influenza B) are 1997/98, 1999/00, 2002/03 and 2003/04 (illustrated in Figure 1).

Estimates for an individual season are not as robust as one would wish. The main purpose of this statement was to respond to the question of whether increases in vaccination coverage in recent years have resulted in a reduction in absenteeism attributable to influenza. Based on currently available data, the answer is that a reduction was not evident. A possible explanation is that a higher overall attack rate due to more than one circulating strain could have off set the benefits of vaccination in this age group (working age adults). Social pressures to stay home if sick may also have contributed.

The paragraph has been revised accordingly.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. General: "Own illness or disability" sounds awkward. Simply "illness or disability" would be preferable, except perhaps for the first mention in line 98.

Author response:

This is the standard language of the LFS, as there is a separate question regarding absences for the care of others.
A description of data related to the care of others was added to the methods section, as we had hoped to model this data as well. "Own illness or disability" has been removed in some instances to improve readability.

2. General: throughout the text, table, figures, and captions, I would use "influenza" rather than "flu" (except where they refer to the other study in the Discussion)

Author response:

Corrected as suggested. The document was check for the use of "flu" rather than "influenza". All remaining uses of "flu" refer to the specific question used in the special LFS question. Substitution of 'influenza' in that case would be misleading as there was no way to confirm whether a 'flu' absence meant 'influenza'.

3. Line 66: "Our" should probably be changed to "Canada's"

Author response: Corrected.

4. Suggest deleting sentence that starts on line 67 – I'm not sure that it is needed.

Author response:

The sentence was revised to indicate that as a result, case ascertainment is a challenge, and hence the need for statistical estimates.

5. Line 102: suggest changing "clinical" to "outpatient"

Author response: Revised.

6. Line 129: missing "for" after "account"; replace "flu year" with "year" or "influenza season"

Author response: Revised. FY stands for Flu Year.

7. Line 165: missing "of" and suggest deleting "their own"

Author response: Corrected.

8. Line 185: should be "loss" rather than "lost"

Author response: Corrected.

9. Line 212: should add reference for this sentence.

Author response: Revised.

10. Line 229: missing "of"

Author response: Corrected.

11. Line 263: references for hospitalizations should be #1 and #16; references for mortality should be #6 and #28

Author response: Reference#1 included.

12. Line 269: suggest "but were" instead of "though"

Author response: Revised.

13. Figure 2: it's not clear which of the two y-axes correspond to which lines.

Author response: The secondary axis was indicated in the legend.

14. Figure 3: why are confidence intervals present for some of the bars but not

others?

Author response: Basically legibility. Caption was revised.

15. Figure 4b: the legend is incorrect (should be absenteeism rate, not % hours lost)

Author response: The label was revised.

16. Figure 6 seems unnecessary – the baselines are already demonstrated in Figure 4.

Author response: Figure 6 was included to highlight the strong difference in seasonal component between the two time series. It could be omitted.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

1. *Line 235: Suggest reporting as a percentage rather than as "5/6th"*

Author response: The LFS has been described in more detail, indicating that each panel consists of 6 sub panels and that households remain in the survey for 6 months, with one of the sub-panels replaced each month. $5/6^{th}$ has been replaced with 5 out of 6 households.

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I declare that I have no competing interests

Reviewer's report
Title: Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey
Version: 1 **Date:** 16 November 2010
Reviewer: Richard Pitman
Reviewer's report:

Major Compulsory Revisions
None

Minor Essential Revisions
Background:
Line 5: It is erroneous to claim that absenteeism is the main economic driver. This depends on the adopted perspective, however, hospitalisations in the elderly is also a very significant driver in economic analyses.

Author response:
As mentioned, it depends on the adopted perspective. Economic costs associated with absenteeism are more likely to be included in studies assessing the cost-effectiveness of workplace vaccination. A recent ESWI bulletin notes that economic costs associated with absenteeism due to influenza have been overlooked: "if you want to maximise economic savings, then target the working population."
http://www.eswi.org/userfiles/files/ESWI_answeringtheinfluenzathreat.pdf

The text has been revised and the above reference included.

Line 12: Remove the word "Our" and replace with "The Canadian".

Author response: Corrected.

Methods:
Data Sources second paragraph line 1: Insert "of" after "The weekly number".

Author response: Corrected.

Statistical Analysis second paragraph: The model used for absenteeism should be clearly defined.

Author response: The dependant variable has been defined in more detail. There were slight differences in the calculations of the annual summary rates for absenteeism and hours lost due to influenza, and the appropriate descriptions have also been included in the methods section.

Results:
Annual Estimates line 8: "lost" should read "loss"

Author response: Corrected.

Discussion:

Second paragraph line 22: Please delete the sentence starting "In a separate analysis...". The discussion is not the place to introduce new results, especially when full results and the associated methods are not described.

Author response: One of the first questions of my co-authors who were actively involved in the study of 'flu' absenteeism using the "4 special LFS questions" was whether differences in the two approaches to estimating absenteeism during the pandemic could be attributed to absences due to care for others. The answer is no, there was no evidence of excess absences during periods with significant influenza activity associated with care for others.

The sentence has been revised and moved to the methods section.

Figure 2: *None of the traces refer specifically to influenza B. The sentence starting "The level of influenza B..." is confusing and should be removed (the point is already covered in the text).*

Author response: The sentence has been removed.

Figure 4a: Change legend to "Influenza-attributed hours lost" and change ordering to be consistent with 4b.

Author response: The legend has been revised.

Discretionary Revisions

In figure captions, try to avoid starting with "In this figure".

Author response: The captions have been revised.

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.
Declaration of competing interests:
I declare that I have no competing interests

Reviewer's report
Title: Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey
Version: 1 **Date:** 25 November 2010
Reviewer: Mark Jit
Reviewer's report:
The authors apply a well established regression method that is used to determine the proportion of respiratory disease outcomes in time series data that are due to a particular organism (eg. influenza) using laboratory reports for that organism as explanatory variables. While there are variations in testing and reporting within and between seasons, in this case, these are taken into account by the use of indicator variables for both month of the year and influenza year. The novelty of the analysis is that it is applied instead to work-related absenteeism rather than disease outcomes or health care resource use. This application of the method appears legitimate, although there is the concern that it is quite a broad outcome measure (compared to eg. hospitalisations for respiratory outcomes).

Major compulsory revisions
1. My main concern is the comment that variables for non-influenza organisms were dropped due to "lack of statistical significance". There could well be a substantial burden of disease in the relevant age groups due to organisms like RSV and S. pneumoniae. Indeed, since the outcome variable (work-related absenteeism) is not specific to respiratory disease outcomes, the explanatory organisms should not be restricted to respiratory organisms – other organisms (e.g. gastrointestinal) also display seasonality and may lead to absenteeism. The authors need to provide more detail in order to justify this absence, including giving details of the variable selection algorithm, the significance cut-off, the organisms considered and (perhaps in an appendix) a graph of the time-series data for laboratory reports of each organism. The concern is that the variables dropped out of the regression simply because the data are aggregated into monthly counts (rather than eg. weekly or daily) and hence it is difficult to be sensitive to seasonal differences in organisms which peak around the same time (e.g. influenza, RSV, S. pneumoniae, norovirus etc.)

Author response:

A description of the model selection criteria has been added to the methods section and the concerns identified above have been discussed more fully throughout the manuscript.

The concern that there are many infectious organisms that cause work-related absenteeism is a valid concern: one that has been explored in earlier work and by many other authors. The short explanation is that there are many organisms, so that the weekly number of sick days (deaths, or hospital admissions, etc.) resulting from the total of all these omitted causes are likely reasonably described by seasonality and trend.

The main concern in this type of modelling is whether the inclusion or exclusion of one exogenous variable, such as RSV, would significantly

alter the parameter estimate or confidence intervals for the influenza parameters, and thereby change the model results.

It is unfortunate that the LFS data are only available monthly. The close association in timing between the week of peak influenza activity and the peak week in respiratory admissions, deaths, etc. over many seasons, along with the variation in timing of influenza epidemics from season-to-season, has supported the conclusion that it is influenza that is responsible for predominant identifiable excess burden in other weekly time series.

Why does influenza stand out? In paediatric respiratory admissions, both RVS and para-influenza were found to account for a larger proportion of the disease burden than influenza. In analyzing this data, proxies for RVS and para-influenza activity could not be ignored and still hope for meaningful estimates of the influenza burden. Influenza stands out because the timing of peak activity varies from year to year and it is often responsible for a relatively large share of the burden compared to other causes with a similar variation in timing. If the latter condition is not met, exogenous variables are required in the model to account for any source of systematic variation (for example RVS in paediatric models, and various holiday effects in hospital admissions). With rhinovirus, corona virus and hMPV recently included in the *FluWatch*, this question can be revisited after a few years of data to test the former hypothesis. (FluWatch historically collected laboratory test results for RVS, adenovirus and parainfluenza as well as influenza A and B.)

2. There are a large number of coefficients in the regression equation (I counted 40) compared to the number of data points (around 150). While this looks necessary in order to take into account both within and between season trends, this means that the model is likely to fit data points well for this reason alone. Hence the results of the analysis need to be interpreted with some caution given that there is no way to independently validate them. This caveat probably needs to be mentioned in the discussion, unless the authors can think of a means of independent validation using a separate data source.

Author response:

Further discussion of over fitting has been included in the methods and discussion sections. Over fitting becomes a problem when parameter estimates start to vary widely and generally loose statistical significance. The use of the number of influenza A positive tests along with separate parameters for each season, in addition to providing a slightly better model fit compared to using the percent of tests positive for influenza A, provided a redundancy that helped illustrate model robustness. The effect of influenza B on workplace absenteeism (and hours lost) could not be estimated separately for each season due to limited statistical power and limited impact of influenza B on workplace absenteeism, so that *fluBpp* was

used as one approach to account for the effects of influenza B in the model without over fitting.

Limited independent validation has been documented in a number of ways: 1) by comparing the results of this model with the results of the 4 special questions; 2) by comparing the estimated absenteeism and hours lost due to influenza for each season (along with the use of separate parameter estimates for each season); 3) by comparing the seasonal estimates of hours lost due to influenza and absenteeism due to influenza (Figure 1); and 4) by using a scale parameter to account for less than optimal model fit, which inflates the corresponding confidence intervals for parameter estimates.

Caution is still advised, as this is a population level study and the manuscript has been reviewed to set a more cautionary tone.

Minor essential revisions

1. *Was the coefficient beta_5 (time off for non-influenza related ILI after the pandemic was announced) significant? I would be concerned if it was – while it is quite plausible that the pandemic situation would cause workers to take precautionary time off, it is difficult to believe that the amount of time was identical throughout the pandemic.*

Author response:

The pandemic indicator variable *Pandemic2009* was included in the model to account for any change in sick leave behaviour once the pandemic was announced that was not related to the level of influenza activity, that is an influenza infection (for example, staying home because of concern that a respiratory infection was due to pandemic influenza even though the employee would have otherwise reported for work and the infection was due to another respiratory virus). The expectation was that beta_5 would be positive due the public health messaging to 'stay home if sick' and because similar increases had been seen in ILI related physician visits (large increases attributed to influenza infections and a smaller increase in the seasonal baseline attributed to non-influenza related ILI). The beta_5 parameter was actually negative in the hours lost model, and not significant in the absenteeism model. The negative impact can be confirmed visually in Figure 4 where the proportion of hours lost is lower in the summer months than it had been for many year. Perhaps employees took less time off in anticipation of possibly needing additional sick days due to the pandemic.

Modifications have been made to the results section and to the discussion.

Also what was used to decide when the pandemic ended?

Author response:

The text was modified to indicate that the pandemic period started in May 2009 and continued until the end of the study period, February 2010. Agreed, behaviour change may not have been constant over this full period. As 3.5% of employed persons indicated that they stayed home from work because of the 'flu' in January in the 4 special LFS questions, February was considered part of the pandemic period.

2. Were all hospitalisations for respiratory causes tested for H1N1 during the pandemic?

Author response:

We will have a better idea of the completeness of testing and reporting once the full hospitalization database for 2009 is released. This question will be the basis of another study though it is too early to comment fully. Some hospitals reported testing all respiratory admissions (personal communication).

3. In Table 1, it's not clear why the % of hours lost attributable to influenza is much greater than the % of hours lost due to own illness or disability – surely the first is a subset of the second.

Author response:

The column titles have been revised to better clarify the different denominators.

The '% of hours lost attributable to influenza' should read 'the proportion of hours lost that were attributable to influenza', i.e. the denominator is hours lost.

The '% of hours lost due to own illness or disability' is hours lost/potential hours worked.

4. In Figure 3, why are the error bars present only for a few characteristics?

Author response:

The CIs were quite broad and differences by employment characteristics were not statistically significant. The CIs for the Urban/Rural split were included to illustrate. The others were not included for legibility. The caption was modified to provide this explanation.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I declare that I have no competing interests