PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Using population-wide administrative and laboratory data to estimate type- and subtype-specific influenza vaccine effectiveness: a surveillance protocol
AUTHORS	Scott, Allison; Buchan, Sarah; Kwong, Jeff; Drews, Steven; Simmonds, Kimberley; Svenson, Lawrence

VERSION 1 – REVIEW

REVIEWER	Kylie Carville Victorian Infectious Diseases Reference Laboratory, The Peter Doherty Institute for Infection and Immunity, Australia
REVIEW RETURNED	15-Mar-2019

GENERAL COMMENTS	This paper describes a useful approach to influenza VE estimation. Some elements of the paper could be explained in more detail. Statistical analysis says 'All covariates will be considered for the adjusted model'. Can this be more specific - will all be included regardless of significance, or will some elimination procedure be used (and what this is)? Will VE estimates be calculated separately for age groups of particular interest ie <5, 65 and over? Will time since vaccination be included in the model, given evidence that the vaccine wanes during the season (6-11% per month, Ferdinands CID 2017), and 3 months elapses between vaccination campaigns and peak activity page 7 lines 129-30. Does the linkage unit do validity checks? Some more details on data sets: Does Alberta have workplace vaccination? If so are these included in the 'mass vaccination clinics' recorded in Imm/ARI? The SESE and MACAR systems - How many diagnosis codes can be entered in these systems? Are there free text fields too that are not extracted for linkage? ie if a patient sees a GP for their diabetes and an ARI, will both be recorded? Or is there a chance diabetes would be the diagnosis code and ARI would only be recorded in free text (which is not available as far as my reading of the protocol). Has validation of these data sources been done? AHCIP: how is socio-economic status determined? It would be beneficial if the authors could check over the paper again for editing - eg acute respiratory infection is defined as ARI on page 8 line 147 but is still spelt out on page 11 lines 214-5.

REVIEWER	Silvio Tafuri

	Aldo Moro University of Bari
REVIEW RETURNED	20-Mar-2019

GENERAL COMMENTS	dear editor, thank you for the opportunity of reviewing this protocol. The field of the research is interesting and actual and I have only few suggestion to improve the protocol. Abstract: Please, clarify in the abstract the setting of the research Article summary: the first bullet point, in the current formulation, is not related to S&L Introduction page 6, line 85: write "disease" instead of "pathogen" line 89: check the rate of flu in Canada... are data presented related to the severe cases? line 90. please, clarify that sentence is referred to 2018/19 season line 95/97: sentence is not clear. Please rewrite line 121: some data about VE evaluated in previous studies could add value to the protocol Laboratory methods must be described in depth. Statistical analysis: please, add the list of covariates considered

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Kylie Carville
Institution and Country: Victorian Infectious Diseases Reference Laboratory, The Peter Doherty Institute for Infection and Immunity, Australia .

1. Statistical analysis says 'All covariates will be considered for the adjusted model'. Can this be more specific - will all be included regardless of significance, or will some elimination procedure be used (and what this is)?
• Response: Yes, all of the covariates are placed in the model regardless of significance. They are being included as they are deemed scientifically important to control for confounding.

2. Will VE estimates be calculated separately for age groups of particular interest ie <5, 65 and over?
• Response: Yes, when we have adequate sample size for a particular season, the VE for those age groups will be estimated.

3. Will time since vaccination be included in the model, given evidence that the vaccine wanes during the season (6-11% per month, Ferdinands CID 2017), and 3 months elapses between vaccination campaigns and peak activity page 7 lines 129-30.

• Response: Thank you very much for pointing this out. We had omitted an element of time from our list of covariates. Month will be included in the model as the rate of exposure to influenza varies significantly over the season. Approximately two-thirds of all individuals who will receive the vaccine in a particular season do so within the first 2 weeks of the public immunization campaign. Other than in pandemic seasons where there is significant public fear around the time of the peak (such as the 2009/10 influenza season), the rate of immunization drops off rapidly after the first two weeks. Given that the start time for the analysis is one month after the start of the public campaign when the vast majority of vaccinations will be complete, and adjusting for both month and time since vaccination would cause over adjustment in the model, we have chosen to adjust only for month.

4. Does the linkage unit do validity checks?
• Response: Deterministic linkage is carried out across the multiple databases. In Alberta, the personal health number (PHN) is a unique identifier used across all provincially funded health care systems. All systems have quality assurance programs in place to ensure that the health record is matched with the correct person. The linkage is then carried out by surveillance analysts in the Ministry of Health. Validation of the PHNs is performed by confirming the sex and date of birth of individuals across the databases.

5. Does Alberta have workplace vaccination? If so are these included in the 'mass vaccination clinics' recorded in Imm/ARI?
• Response: Alberta does have some workplace vaccination clinics and vaccination clinics at universities and other post-secondary educational institutions. While all workplace and educational institutional vaccination clinics are scheduled for eventual inclusion in Imm/ARI, the majority are not at this time. The number of doses of vaccine distributed to these groups is less than 5% of vaccination in Alberta.

6. The SESE and MACAR systems - How many diagnosis codes can be entered in these systems? Are there free text fields too that are not extracted for linkage? ie if a patient sees a GP for their diabetes and an ARI, will both be recorded? Or is there a chance diabetes would be the diagnosis code and ARI would only be recorded in free text (which is not available as far as my reading of the protocol).
• Response: Up to 25 diagnostic codes are included in the MACAR system and up to 3 diagnostic codes are included in the SESE system. There are no free text fields in either system. If a patient is admitted to a hospital, emergency room or outpatient clinic there are specialized abstractors that code all of the diagnoses given to a patient and include it in the MACAR system, therefore the diabetic patient with ARI you referenced in your question would get both codes. GPs are required to provide at least one diagnostic code for each patient encounter, but are not required to provide more than one. How the patient would be coded would vary depending on the diligence of the GP and which concern was considered the main reason for the visit.

7. Has validation of these data sources been done?
• Response: Yes, there has been significant work to establish the validity of these administrative datasets in Alberta. Please see the following references. A note has been placed in the methods.
i. Shiff NJ, Jama S, Boden C, Lix LM. Validation of administrative health data for the pediatric population: a scoping review. BMC health services research. 2014 Dec;14(1):236.
ii. Quan H, Smith M, Bartlett-Esquilant G, Johansen H, Tu K, Lix L. Mining administrative health databases to advance medical science: geographical considerations and untapped potential in Canada. Canadian Journal of Cardiology. 2012 Mar 1;28(2):152-4.
iii. Hinds A, Lix LM, Smith M, Quan H, Sanmartin C. Quality of administrative health databases in Canada: a scoping review. Canadian Journal of Public Health. 2016 Jan 1;107(1):e56-61.

8. AHCIP: how is socio-economic status determined?
• Response: Socio-economic status is derived from census dissemination area income quintiles using postal code. This has been updated in the methods.

9. It would be beneficial if the authors could check over the paper again for editing - eg acute respiratory infection is defined as ARI on page 8 line 147 but is still spelt out on page 11 lines 214-5.
• Response: Thank you. We have read the paper over and corrected the errors.

Reviewer: 2
Reviewer Name: Silvio Tafuri

Institution and Country: Aldo Moro University of Bari
Please state any competing interests or state ‘None declared’: None to declare

1. Abstract: Please, clarify in the abstract the setting of the research
• Response: We have clarified in the abstract that this is a collaborative effort by the Ministry of Health in Alberta and the Provincial Public Health Laboratory in Alberta.

2. Article summary: the first bullet point, in the current formulation, is not related to S&L
• Response: We have rewritten this statement.

3. Introduction page 6, line 85: write "disease" instead of "pathogen"
• Response: Completed.

4. line 89: check the rate of flu in Canada... are data presented related to the severe cases?
• Response: This is the rate of laboratory-confirmed cases. We have revised the sentence to better reflect this.

5. line 90. please, clarify that sentence is referred to 2018/19 season
• Response: This number is an overall average of what is normally seen in a typical season. We have revised the sentence for clarity.

6. line 95/97: sentence is not clear. Please rewrite
• Response: Completed.

7. line 121: some data about VE evaluated in previous studies could add value to the protocol
• Response: There has been one published estimate of Alberta-specific vaccine effectiveness using a sentinel surveillance system (Skowronski et al 2017. Interim estimates of 2016/17 vaccine effectiveness against influenza A(H3N2), Canada, January 2017. Euro Surveill, 2017; 22(6):pii:3060.). However, because of the small sample size the confidence interval was large, ranging from 8% to 72%. Some of us have prepared a manuscript (currently under review) that indicates that VE estimates generated using linked health administrative and laboratory data in the province Ontario are comparable to previously published estimates.

8. Laboratory methods must be described in depth.
• Additional information and one additional reference was provided.

9. Statistical analysis: please, add the list of covariates considered
• Response: Completed.

VERSION 2 – REVIEW

REVIEWER	Kylie Carville VIDRL at the Peter Doherty Institute for Infection and Immunity, Australia
REVIEW RETURNED	19-May-2019

GENERAL COMMENTS	Thanks for the chance to review. I have one key question, apologies if I have missed this somewhere as it seems fundamental. The study is promoted as being 'near real time'. I presume when this article was first submitted the 2018/19 data

	collection was in the future (hence originally described as the upcoming season). It would be nice to know if the linkage/analysis was done throughout the 2018/19 season, given the years of historical data can be linked once after the fact. I presume the data sets are (or can be) linked on a regular, or episodic, basis during the season to enable near real time. How frequently is this?

VERSION 2 – AUTHOR RESPONSE

The methods were tested in the 2018/19 season but were not performed in real time. Implementation of the real-time analyses will occur in the upcoming season. One of the strengths of the method is that it can be implemented quickly whenever required. So at times when there is high concern (such as at the start of the influenza season or in a pandemic) it can be implemented weekly or biweekly, while at times of low concern it may be implemented once or twice a month. A likely schedule for the upcoming season (assuming it is a typical influenza season) would be biweekly until the peak of the influenza season and monthly thereafter. The date range has been updated in this version.

VERSION 3 – REVIEW

REVIEWER	Kylie Carville VIDRL at the Peter Doherty Institute for Infection and Immunity
REVIEW RETURNED	03-Jul-2019

GENERAL COMMENTS	I would still appreciate some clarity on what 'near real time' is, or what it is expected to be, or a statement that this will be determined by the study protocol. Adding an extra year to the protocol to ensure it is prospective, not retrospective, doesn't address this query. I don't think this is sufficient to block publication, but I do think that clarity in this matter would be appreciated by readers.

VERSION 3 – AUTHOR RESPONSE

My apologies if our use of the term near-real time was vague. We have revised the paper to indicate that the VE will be estimated every two weeks prior to the peak of the season and estimated every month thereafter. Does that address your concern? Please feel free to contact me at 1-587-635-0613 if it does not.