

RESEARCH

Open Access



# Quantifying the annual incidence and underestimation of seasonal influenza: A modelling approach

Zachary McCarthy<sup>1,2,3,4</sup>, Safia Athar<sup>1,2,3,4</sup>, Mahnaz Alavinejad<sup>1,2,3,4</sup>, Christopher Chow<sup>1,2,3,4</sup>, Iain Moyles<sup>1</sup>, Kyeongah Nah<sup>1,2,3,4</sup>, Jude D. Kong<sup>1,2,3,4</sup>, Nishant Agrawal<sup>5</sup>, Ahmed Jaber<sup>6</sup>, Laura Keane<sup>1</sup>, Sam Liu<sup>7</sup>, Myles Nahirniak<sup>6</sup>, Danielle St Jean<sup>6</sup>, Razvan Romanescu<sup>8</sup>, Jessica Stockdale<sup>6</sup>, Bruce T. Seet<sup>9,10</sup>, Laurent Coudeville<sup>11</sup>, Edward Thommes<sup>2,6,9</sup>, Anne-Frieda Taurel<sup>9</sup>, Jason Lee<sup>9</sup>, Thomas Shin<sup>9</sup>, Julien Arino<sup>12</sup>, Jane Heffernan<sup>1,2,3,4</sup>, Ayman Chit<sup>13,14</sup> and Jianhong Wu<sup>1,2,3,4\*</sup>

## Abstract

**Background:** Seasonal influenza poses a significant public health and economic burden, associated with the outcome of infection and resulting complications. The true burden of the disease is difficult to capture due to the wide range of presentation, from asymptomatic cases to non-respiratory complications such as cardiovascular events, and its seasonal variability. An understanding of the magnitude of the true annual incidence of influenza is important to support prevention and control policy development and to evaluate the impact of preventative measures such as vaccination.

**Methods:** We use a dynamic disease transmission model, laboratory-confirmed influenza surveillance data, and randomized-controlled trial (RCT) data to quantify the underestimation factor, expansion factor, and symptomatic influenza illnesses in the US and Canada during the 2011-2012 and 2012-2013 influenza seasons.

**Results:** Based on 2 case definitions, we estimate between 0.42 – 3.2% and 0.33 – 1.2% of symptomatic influenza illnesses were laboratory-confirmed in Canada during the 2011-2012 and 2012-2013 seasons, respectively. In the US, we estimate between 0.08 – 0.61% and 0.07 – 0.33% of symptomatic influenza illnesses were laboratory-confirmed in the 2011-2012 and 2012-2013 seasons, respectively. We estimated the symptomatic influenza illnesses in Canada to be 0.32 – 2.4 million in 2011-2012 and 1.8 – 8.2 million in 2012-2013. In the US, we estimate the number of symptomatic influenza illnesses to be 4.4 – 34 million in 2011-2012 and 23 – 102 million in 2012-2013.

**Conclusions:** We illustrate that monitoring a representative group within a population may aid in effectively modelling the transmission of infectious diseases such as influenza. In particular, the utilization of RCTs in models may enhance the accuracy of epidemiological parameter estimation.

**Keywords:** Mathematical modelling; influenza; vaccine; evidence synthesis

\*Correspondence: [wujh@yorku.ca](mailto:wujh@yorku.ca)

<sup>1</sup>Department of Mathematics and Statistics, York University, M3J 1P3 Toronto, Canada

<sup>2</sup>Laboratory for Industrial and Applied Mathematics, York University, M3J 1P3 Toronto, ON, Canada

<sup>3</sup>Centre for Disease Modelling, York University, M3J 1P3 Toronto, ON, Canada

<sup>4</sup>Fields-CQAM Mathematics for Public Health Laboratory, York University, M3J 1P3 Toronto, ON, Canada

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

The exact number of cases of a disease is complex to capture. Different methods can be used, from epidemiological studies to disease surveillance systems. While data collected routinely for surveillance purposes have the advantage of being readily accessible over a long period of time, they are subject to underestimation. Underestimation is a combination of under-reporting (failure to capture cases that seek care due to underdiagnoses or under-notifications) and under-ascertainment (failure to seek health care) [1]. Symptomatic individuals who seek medical care but are misdiagnosed due to an atypical presentation which does not fit the case definition or to the lack of sensitivity of the laboratory test (under-diagnosis) and/or for which administrative steps may not be taken at the physician's office to report the case contribute to under-reporting [1]. Also, infected individuals may be asymptomatic or with a mild form of the disease and may not seek healthcare, leading to under-ascertainment. Mathematical modelling may play a role in quantifying the effects of factors contributing to underestimation to assess the true number of influenza illnesses, ultimately to assist in policy development and to evaluate the impact of influenza vaccination.

Passive influenza surveillance systems are not designed to capture all illnesses; however, surveillance data has been utilized to assess under-reporting, underestimation and incompleteness. Statistical modelling has been utilized to quantify under-reporting and underestimation of influenza-associated hospitalizations, morbidity and mortality in the United States (US) and Canada [2–12]. However, assessing the underestimation associated with the true number of influenza illnesses has received less attention. Influenza surveillance data has been utilized to assess underestimation and estimate symptomatic influenza illnesses using multiplier methods [13, 14], Bayesian evidence synthesis [15], and dynamic models using ordinary differential equations [16]. In one study, assessments of underestimation were shown to be unreliable when utilizing only virologic surveillance data [16]. Similarly, a noted contributor of error of symptomatic illness estimates was the quality of sentinel influenza-like illness (ILI) data, which is subject to both overestimation and underestimation [15]. A method to assess underestimation of influenza illnesses utilizing a reliable data source, such as active surveillance, may enhance the ability to quantify the true underestimation and the number of influenza illnesses. In the present study, we develop a framework to utilize data from a closely monitored group whose epidemiological status is subject to minimal uncertainty. In particular, we utilize a randomized-controlled trial (RCT) designed to minimize the effects of underestimation and completely capture participant's illnesses through active surveillance and systematic testing [17].

A recent vaccine RCT, which assessed the efficacy of a high-dose (HD) influenza vaccine compared to a standard-dose (SD) influenza vaccine, presents an opportunity to explore a new method using mathematical modelling to correct for underestimation in national virological influenza surveillance, ultimately to infer the number of symptomatic influenza illnesses. In particular, we establish a parameter estimation technique based on a mechanistic disease transmission model to assess the underestimation and symptomatic illnesses in the US and Canada during the 2011-2012 and 2012-2013 influenza seasons.

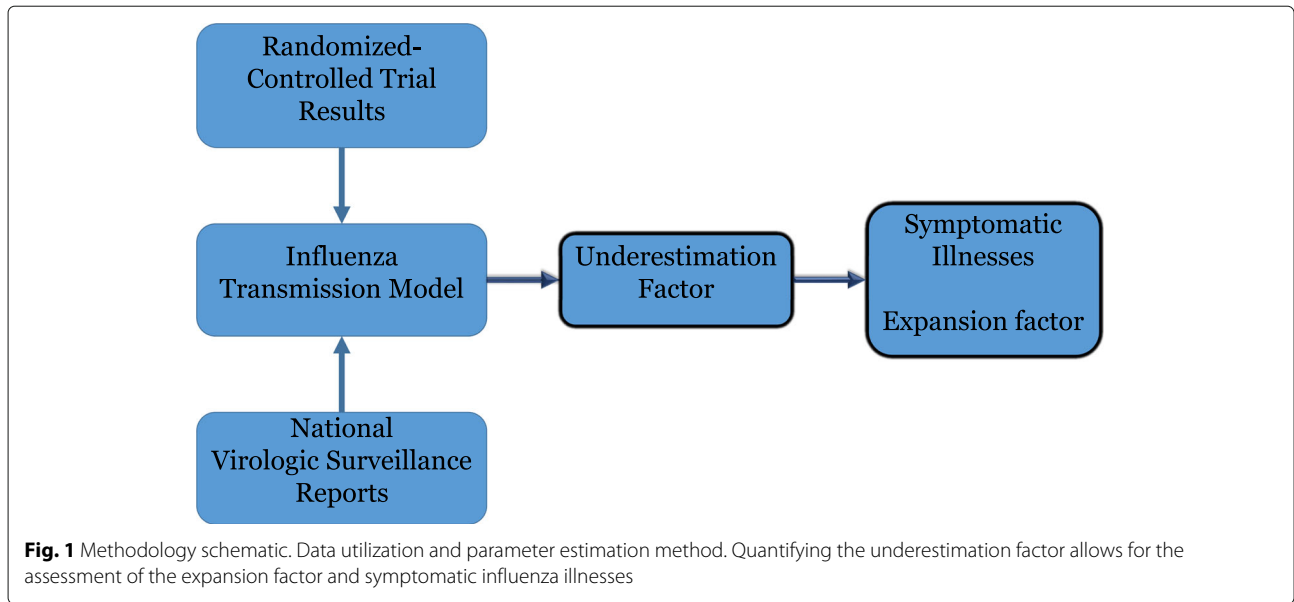
## Methods

### Data sources

The present modelling study integrates multiple sources of data to generate estimates for the underestimation factor, disease transmission rate, expansion factor, and symptomatic influenza illnesses (Fig. 1). The laboratory-confirmed influenza cases during the 2011-2012 and 2012-2013 seasons are included in the final seasonal surveillance reports from FluWatch and FluView in Canada and the US, respectively [18, 19]. Additionally, we utilize the US and Canadian census profiles [20–23].

A recent RCT assessed the efficacy of a HD influenza vaccine compared to a SD influenza vaccine among 14,500 and 17,489 participants during the 2011-2012 and 2012-2013 seasons, respectively [17]. We utilize the laboratory-confirmed influenza counts among participants receiving SD and HD influenza vaccines in this RCT, which have been provided in the Supplementary Appendix of a prior publication [17]. The number of laboratory-confirmed influenza cases within the RCT are provided for three different case definitions [17]. In the present study, we utilize the laboratory-confirmed influenza cases associated with the least and most restrictive case definitions in the RCT. We utilize data associated with the least restrictive case definition, respiratory illness (RI), to provide estimates representing the true underestimation of symptomatic influenza illnesses. We utilize a more restrictive case definition, modified CDC-defined ILI, which is closer to the case definition used by surveillance systems in the US and Canada, to provide estimates representing underestimation of symptomatic influenza illnesses which were captured by virologic surveillance. RI was defined as the occurrence of one or more of the following: sneezing, nasal congestion or rhinorrhea, sore throat, cough, sputum production, wheezing, or difficulty breathing and modified CDC-defined ILI was defined as RI with cough or sore throat, concurrent with a temperature above 37.2° C [17].

The SD influenza vaccine effectiveness and coverage over the 2011-2012 and 2012-2013 influenza seasons has been established in the US and Canada [24–27]. Also, the



HD influenza vaccine coverage estimates have been made available in the US [28]. In Canada, the HD influenza vaccine was not yet licensed for use during 2011-2012 and 2012-2013 seasons.

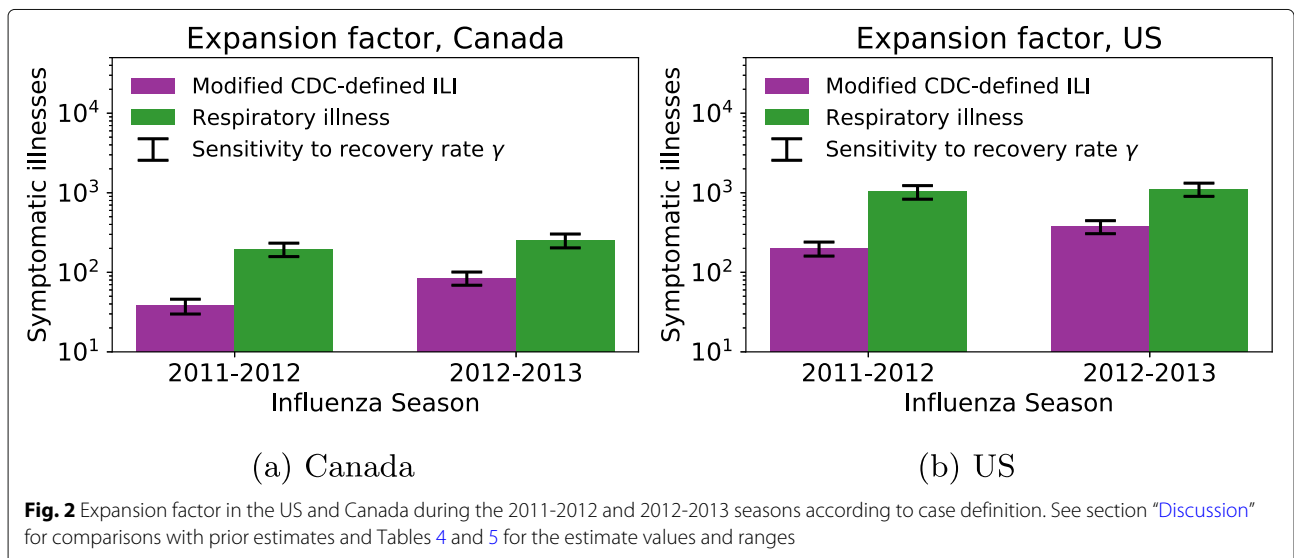
**Mathematical modelling**

We develop two compartmental models consisting of vaccinated and infected individuals for the 1) seniors aged 65+ who participated in the RCT and 2) the general population. Individuals are assumed to be vaccinated by the beginning of the influenza season, so there is no in-flow to the vaccinated compartment during the season. Since we study the dynamics of influenza within one season, we ignore the demographic dynamics (e.g. birth and death). As the size of the population for participants

of the RCT is small relative to the total population, we ignore the influenza transmission from these participants to the general community; however, we assume the same transmission rate among the RCT participants. Also, we assume that the age groups 0-64 and 65+ are homogeneously mixed. In addition to homogeneous mixing by age group, we assume spatial homogeneity in transmission. We assign units of symptomatic influenza illness to the infected compartments of the following models for dimensional consistency with data.

**Modelling influenza infections among RCT participants**

We develop a compartmental model consisting of vaccinated and infected individuals for the seniors aged 65+ who participated in the RCT [17]. We make use of the



following notation for model variables and parameters; the subscript “+” refers to compartments of individuals aged 65+ and subscript “-” refers to individuals aged 0-64. The vaccination status of the compartment will also appear in the subscript as SD (standard-dose) or HD (high-dose), if applicable. Lastly, let the superscript “O” denote individuals outside of the RCT and the superscript “C” denote participants within the RCT. With these assumptions and notations set, we formulate the following system of ordinary differential equations, where  $V_{+,SD}^C$  represents RCT participants vaccinated given the standard-dose vaccine,  $V_{+,HD}^C$  represents RCT participants vaccinated given the high-dose vaccine,  $I_{+,HD}^C$  represents infected RCT participants with high-dose vaccine, and  $I_{+,SD}^C$  represents the infected RCT participants given standard-dose vaccine. These model variables are described in Table 1. Submodel (1) is equipped with the following initial conditions to replicate the RCT conditions preceding each season:  $V_{+,SD}^C(0) = V_{SD0}$ ,  $V_{+,HD}^C(0) = V_{HD0}$ ,  $I_{+,SD}^C(0) = 0$ ,  $I_{+,HD}^C(0) = 0$  [17].

$$\begin{aligned}
 \dot{V}_{+,SD}^C &= -\epsilon_{+,SD} V_{+,SD}^C \beta (I_{+,SD}^C + I_{+,HD}^C + J) \\
 \dot{V}_{+,HD}^C &= -\epsilon_{+,HD} V_{+,HD}^C \beta (I_{+,SD}^C + I_{+,HD}^C + J) \\
 \dot{I}_{+,SD}^C &= \epsilon_{+,SD} V_{+,SD}^C \beta (I_{+,SD}^C + I_{+,HD}^C + J) - \gamma I_{+,SD}^C \\
 \dot{I}_{+,HD}^C &= \epsilon_{+,HD} V_{+,HD}^C \beta (I_{+,SD}^C + I_{+,HD}^C + J) - \gamma I_{+,HD}^C.
 \end{aligned} \tag{1}$$

Parameters  $\epsilon_{+,SD}$  and  $\epsilon_{+,HD}$  denote the vaccine-modified susceptibilities corresponding to the standard-dose and high-dose influenza vaccine efficacy among seniors, respectively,  $\beta$  is the transmission rate and  $\gamma$  is the influenza recovery rate. Specifically,  $\beta = b/N$  where  $b$  is the daily effective contact rate and  $N$  is the total population size. The vaccine-modified susceptibilities  $\epsilon$  represent a multiplier for the reduced infection rate of vaccinated individuals who are susceptible to influenza infection. Lastly,  $J$  represents the total infected individuals in the population from model (2),  $J = I_-^O + I_+^O$ .

**Country-wide influenza transmission model**

We subdivide the general community (i.e., the entire US or Canada) into two age groups: seniors aged 65+ and

non-seniors aged 0-64, as well as vaccination status (SD or HD). The model variables and notations for model (2) are displayed in Table 2 and described in more detail below.

$$\begin{aligned}
 \dot{S}_-^O &= -S_-^O \beta (I_-^O + I_+^O) \\
 \dot{V}_{-,SD}^O &= -\epsilon_{-,SD} V_{-,SD}^O \beta (I_-^O + I_+^O) \\
 \dot{I}_-^O &= S_-^O \beta (I_-^O + I_+^O) + \epsilon_{-,SD} V_{-,SD}^O \beta (I_-^O + I_+^O) - \gamma I_-^O \\
 \dot{S}_+^O &= -S_+^O \beta (I_-^O + I_+^O) \\
 \dot{V}_{+,SD}^O &= -\epsilon_{+,SD} V_{+,SD}^O \beta (I_-^O + I_+^O) \\
 \dot{V}_{+,HD}^O &= -\epsilon_{+,HD} V_{+,HD}^O \beta (I_-^O + I_+^O) \\
 \dot{I}_+^O &= S_+^O \beta (I_-^O + I_+^O) + \epsilon_{+,HD} V_{+,HD}^O \beta (I_-^O + I_+^O) \\
 &\quad + \epsilon_{+,SD} V_{+,SD}^O \beta (I_-^O + I_+^O) - \gamma I_+^O
 \end{aligned} \tag{2}$$

As identical with model (1),  $\beta = b/N$  where  $b$  is the daily effective contact rate and  $N$  is the total population size. Model (2) is equipped with the following initial conditions: the influenza season begins with the initial vaccinated population  $V_{-,SD}^O(0) = \rho_{-,SD}^O N_-$ ,  $V_{+,SD}^O(0) = \rho_{+,SD}^O N_+$ ,  $V_{+,HD}^O(0) = \rho_{+,HD}^O N_+$ , where  $\rho_{-,SD}^O$ ,  $\rho_{+,SD}^O$ ,  $\rho_{+,HD}^O$  are the vaccine coverages for the population aged 0-64 and population aged 65 or over (either vaccinated with standard-dose or high-dose) in each year and country. Similarly, the initial susceptible populations are:  $S_-^O(0) = (1 - \rho_{-,SD}^O) N_-$ ,  $S_+^O(0) = (1 - \rho_{+,HD}^O - \rho_{+,SD}^O) N_+$ ; there are no infections preceding the epidemic, so  $I_-^O(0) = I_+^O(0) = 0$ . The values for these parameters embedded in the initial conditions are shown in Table 3 and estimated in “Parameter estimation” section.

**Table 1** Model variables and descriptions for RCT influenza transmission submodel (1)

Model Variable	Description
$V_{+,SD}^C$	RCT participants vaccinated with SD
$V_{+,HD}^C$	RCT participants vaccinated with HD
$I_{+,SD}^C$	Infected RCT participants vaccinated with SD
$I_{+,HD}^C$	Infected RCT participants vaccinated with HD
$J$	Total, country-wide infected population

**Table 2** Model variables and descriptions for the general community (i.e., individuals outside of the RCT) influenza transmission model (2)

Model variable	Description
$S_-^O$	Population aged 0-64 susceptible to influenza infection
$S_+^O$	Population aged 65+ susceptible to influenza infection
$V_{-,SD}^O$	Vaccinated population aged 0-64 (Standard-dose)
$V_{+,SD}^O$	Vaccinated population aged 65+ (Standard-dose)
$V_{+,HD}^O$	Vaccinated population aged 65+ (High-dose)
$I_-^O$	Population aged 0-64 infected with influenza
$I_+^O$	Population aged 65+ infected with influenza

**Table 3** Parameters used for quantifying the influenza underestimation factor and transmission rate in the US and Canada for years 2011-2012 and 2012-2013, respectively

Parameter	Description	2011-12	2012-13	References
<b>Canada</b>				
$N$	Total population	34,339,328	34,714,222	[20], [21]
$N_-$	Number of community members (those outside RCT) aged 0-64	29,384,093	29,559,285	[20], [21]
$N_+$	Number of seniors in the community aged 65+	4,955,235	5,154,937	[20], [21]
$\hat{R}$	Laboratory-confirmed influenza infections	10,279	26,335	[19]
$\rho_{-,SD}^O$	Vaccine coverage among non-seniors	0.28	0.37	[29]
$\rho_{+,SD}^O$	SD Vaccine coverage among seniors	0.53	0.65	[29]
$\rho_{+,HD}^O$	HD Vaccine coverage among seniors	0	0	[30]
$\epsilon_{-,SD}$	Vaccine-modified susceptibility, general community ages 0-64	0.45	0.49	[25], [24]
$\epsilon_{+,SD}$	Vaccine-modified susceptibility for SD, seniors aged 65+	0.42	0.53	[25], [24]
<b>United States</b>				
$N$	Total population	345,288,000	350,333,000	[22], [23]
$N_-$	Number of community members 0-64	306,110,000	308,827,000	[22], [23]
$N_+$	Number of seniors in the community 65+	39,178,000	41,506,000	[22], [23]
$\hat{R}$	Laboratory-confirmed influenza infections	27,012	75,772	[18]
$\rho_{-,SD}^O$	Vaccine coverage among non-seniors	0.28	0.31	[31], [31]
$\rho_{+,SD}^O$	SD Vaccine coverage among seniors	0.649	0.662	[32], [31]
$\rho_{+,HD}^O$	HD Vaccine coverage among seniors	0.058	0.058	[28]
$\epsilon_{-,SD}$	Vaccine-modified susceptibility, general community ages 0-64	0.53	0.51	[26], [27]
$\epsilon_{+,SD}$	Vaccine-modified susceptibility for SD, seniors aged 65+	0.57	0.71	[26], [33]
$\epsilon_{+,HD}$	Vaccine-modified susceptibility for HD, seniors aged 65+	0.29	0.56	Estimated
<b>Comparative vaccine randomized-controlled trial [17]</b>				
$N_{tr}$	Number of RCT participants	14,497	17,486	[17]
$V_{HD0}$	RCT participants given HD influenza vaccine	7253	8737	[17]
$V_{SD0}$	RCT participants given SD influenza vaccine	7244	8749	[17]
Modified CDC-defined ILI case definition				
$V_{+,HD}^C(\infty)$	SD-vaccinated RCT participants who avoided infection	7233	8639	[17]
$V_{+,SD}^C(\infty)$	HD-vaccinated RCT participants who avoided infection	7243	8651	[17]
Respiratory illness case definition				
$V_{+,HD}^C(\infty)$	SD-vaccinated RCT participants who avoided infection	7187	8419	[17]
$V_{+,SD}^C(\infty)$	HD-vaccinated RCT participants who avoided infection	7196	8468	[17]
<b>Parameter for models (1) and (2)</b>				
$\gamma$	Influenza recovery rate	$\frac{1}{38}(\frac{1}{46} - \frac{1}{31})$	$\frac{1}{38}(\frac{1}{46} - \frac{1}{31})$	[34]

**Disease transmission and burden estimates**

The approach for quantifying the underestimation factor and transmission rate utilizes the modelling framework developed in “[Mathematical modelling](#)” section. The mathematical details are presented in Appendix A. We assume that, since the RCT participants were actively monitored in the community, i.e. instructed to contact their study site if they had any respiratory symptoms and

were, in addition, contacted weekly or bi-weekly by the site, that there was no underestimation of influenza infection within the RCT. The key idea is captured in the relationship  $pJ = R$ ; a percent  $p$  (which denotes the underestimation factor) of symptomatic influenza cases  $J$  yields the laboratory-confirmed influenza cases  $R$ . We use this relationship  $pJ = R$ , the final epidemic size relationships for submodel (1) and model (2), and structure of



the SVIR model to derive a tractable system of nonlinear equations. The representative equation derived from submodel (1) captures information from the RCT, while the representative equation derived from model (2) captures information from the general population. Specifically, we derive a nonlinear system of the form

$$\begin{aligned} f(\beta, p) &= 0 \\ g(\beta, p) &= 0, \end{aligned} \tag{3}$$

where

$$\begin{aligned} f(\beta, p) &= V_{+,SD}^C(\infty) \\ &\quad - V_{+,SD}^C(0)e^{\frac{\epsilon_{+,SD}\beta}{\gamma}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R})} \\ \text{and} \\ g(\beta, p) &= e^{-(\beta p^{-1}\hat{R})} (S_{-,SD}^O(0) + S_{+,SD}^O(0)) + V_{-,SD}^O(0)e^{-(\beta\epsilon_{-,SD}p^{-1}\hat{R})} \\ &\quad + V_{+,SD}^O(0) \left( e^{-(\beta\epsilon_{+,SD}p^{-1}\hat{R})} \right) + V_{+,HD}^O(0)e^{-(\beta\epsilon_{+,HD}p^{-1}\hat{R})} \\ &\quad - (S_{-,SD}^O(0) + S_{+,SD}^O(0) + V_{-,SD}^O(0) + V_{+,SD}^O(0) + V_{+,HD}^O(0)) + \gamma p^{-1}\hat{R}, \end{aligned} \tag{4}$$

and  $\hat{R}$  is the laboratory-confirmed cases in an influenza season captured in national virological surveillance. To solve this system of nonlinear equations (4) for  $p$  and  $\beta$ , we use Matlab's *fsolve* function in Matlab R2016a.

**Expansion factor and symptomatic illnesses**

The expansion factor,  $E$ , is defined in this study as the number of symptomatic influenza illnesses per laboratory-confirmed infection. In terms of the underestimation factor, the expansion factor is its multiplicative inverse. Hence, we compute the expansion factor by finding the multiplicative inverse of  $p$ , that is  $E = p^{-1}$ .

The number of estimated symptomatic influenza illnesses is  $E\hat{R}$  where  $\hat{R}$  is the total laboratory-confirmed influenza cases from national surveillance during an influenza season.

The basic reproduction number is the average number of secondary cases produced by one infected individual introduced into a population of susceptible individuals. We determine the basic reproduction number of model (2) using the next generation method [35].

**Parameter estimation**

We determine estimates for parameters associated with submodel (1) and model (2) utilizing data in “Data sources” section. We provide an outline of the methods in the main text and more complete calculations and explanations are included in Appendix B.

*Estimating vaccine-modified susceptibility  $\epsilon_{-,SD}$ ,  $\epsilon_{+,SD}$  and  $\epsilon_{+,HD}$ :* Recall the vaccine-modified susceptibility  $\epsilon$  captures the protection added from vaccination among vaccinated individuals against influenza infection. Here we outline the method for estimating parameter values

$\epsilon_{+,SD}$ ,  $\epsilon_{+,HD}$  and  $\epsilon_{-,SD}$  for influenza seasons 2011-2012 and 2012-2013. This process integrates model analysis and prior vaccine effectiveness (VE) studies. We make use of a relationship between vaccine-modified susceptibility and VE [36].

To infer  $\epsilon_{-,SD}$  and  $\epsilon_{+,SD}$  we use prior estimates of vaccine effectiveness (VE) against influenza. Specifically, we relate these VE estimates to vaccine-modified susceptibility  $\epsilon$  with the relationship  $\epsilon = 1 - VE$  [36]. Remaining is to find  $\epsilon_{+,HD}$ , which we use  $\epsilon_{+,SD}$  and the ratio  $\epsilon_{+,HD}/\epsilon_{+,SD}$ . We use submodel (1) to determine the ratio  $\epsilon_{+,HD}/\epsilon_{+,SD}$ . Specifically, we find  $\epsilon_{+,HD}/\epsilon_{+,SD}$  by dividing the first two equations in submodel (1), which yields  $\frac{\dot{V}_{+,HD}^C}{V_{+,SD}^C} = \frac{\epsilon_{+,HD} V_{+,HD}^C}{\epsilon_{+,SD} V_{+,SD}^C}$ . Finally, we use separation of variables to find the ratio in terms of known RCT outcomes [17]:

$$\frac{\epsilon_{+,HD}}{\epsilon_{+,SD}} = \frac{\log(V_{+,HD}^C(\infty)/V_{+,HD}^C(0))}{\log(V_{+,SD}^C(\infty)/V_{+,SD}^C(0))}. \tag{5}$$

The quantities  $V_{+,SD}^C(\infty)$  and  $V_{+,HD}^C(\infty)$  are the limit values of the state variables  $V_{+,SD}^C$  and  $V_{+,HD}^C$ , respectively. Finally, we use estimates of now known quantities;  $\epsilon_{+,SD}$  from VE studies and the ratio  $\frac{\epsilon_{+,HD}}{\epsilon_{+,SD}}$  from the RCT to estimate  $\epsilon_{+,HD}$  [17].

*Initial susceptible and vaccinated populations:* To inform the initial conditions for model (2) in the years 2011-2012 and 2012-2013, we use population sizes given by the US and Canadian census programs [20–23]. The population size and the estimated vaccine coverage in each country then gives us the susceptible and vaccinated population initial conditions. The values and descriptions of these parameters embedded in the initial conditions;  $N_{-,SD}$ ,  $N_{+,SD}$ ,  $\rho_{-,SD}^O$ ,  $\rho_{+,SD}^O$ , and  $\rho_{+,HD}^O$ , are listed in Table 3. Similarly, the initial conditions for submodel (1) are given in Table 3.

*Influenza recovery rate  $\gamma$ :* We inform the recovery rate  $\gamma$  using the infectious period of influenza, which has been estimated to be 3.8 days with a 95% confidence interval (CI) of 3.1 - 4.6 days [34]. The recovery rate  $\gamma$  is then the inverse of the mean sojourn time in the infectious compartment; hence, we consider  $\gamma = \frac{1}{3.8} \text{ day}^{-1}$  as a baseline. We utilize the bounds on the CI of the estimated infectious period for sensitivity analysis and vary  $\gamma$  from  $\frac{1}{3.1} \text{ day}^{-1}$ .

We utilize the laboratory-confirmed influenza counts reported in the RCT according to two case definitions to inform  $V_{+,HD}^C(\infty)$  and  $V_{+,SD}^C(\infty)$ . We use 1) laboratory-confirmed cases associated with modified CDC-defined ILI and 2) laboratory-confirmed cases associated with respiratory illness (RI) [17] (Table 3). Results generated for each case definition has an interpretation and is left for the Discussion.

**Well-posedness**

To ensure that the parameter estimation process for obtaining  $p$  and  $\beta$  will yield biologically relevant results, we study the well-posedness of the inverse problem outlined in “[Disease transmission and burden estimates](#)” section. The solution  $p$  and  $\beta$  to system (4) is unique and positive, i.e. the problem is well-posed. From  $f(\beta, p) = 0$ , we note that  $\beta$  is related to  $p$  by the following equation

$$\beta = -\ln \frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} \cdot \frac{\gamma}{\epsilon_{+,SD} V_{+,SD}^C(0) \left(1 - \frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)}\right) + V_{+,HD}^C(0) \left(1 - \frac{V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)}\right) + p^{-1} \gamma \hat{R}}$$

Substituting this expression for  $\beta$  into  $g(\beta, p)$  yields the following equation of  $p$

$$\begin{aligned} & (S_{-}^O(0) + S_{+}^O(0) + V_{-,SD}^O(0) + V_{+,SD}^O(0) + V_{+,HD}^O(0)) - \\ & (S_{-}^O(0) + S_{+}^O(0)) \left(\frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)}\right)^{U(p)} - V_{-,SD}^O(0) \left(\frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)}\right)^{U(p)} \\ & - V_{+,SD}^O(0) \left(\frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)}\right)^{\epsilon_{+,SD} U(p)} - V_{+,HD}^O(0) \left(\frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)}\right)^{\epsilon_{+,HD} U(p)} \\ & = \gamma p^{-1} \hat{R} \end{aligned} \tag{6}$$

where

$$U(p) = \frac{\gamma \hat{R} p^{-1}}{\epsilon_{+,SD} \left( \left( V_{+,SD}^C(0) \left( 1 - \frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} \right) + V_{+,HD}^C(0) \left( 1 - \frac{V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)} \right) + p^{-1} \gamma \hat{R} \right)}$$

Let  $\Phi(p)$  denote the left hand side of equation (6). Note that  $\Phi$  is a monotone-decreasing function of  $p$  with  $\Phi(0) > 0$  and  $\lim_{p \rightarrow \infty} \Phi(p) = 0$ . On the other hand, the right hand side of the equation is a monotone-decreasing function of  $p$  with  $\lim_{p \rightarrow 0} \gamma \hat{R} p^{-1} = \infty$  and  $\lim_{p \rightarrow \infty} \gamma \hat{R} p^{-1} = 0$ . Therefore, there exists a unique solution  $p \in (0, 1]$  if  $\Phi(1) > \gamma \hat{R}$ .

From the parameter sets estimated in “[Parameter estimation](#)” section,  $\Phi(1) > \gamma \hat{R}$ , hence we have a unique solution of System (3) corresponding to each context-specific parameter set. In other words, according to this specific US/Canadian demographic information, vaccine-specific parameters, influenza surveillance reports, and RCT study results, there is a single underestimation factors  $p$  (which we ensure is logically between 0 and 1) and transmission rate  $\beta$  which satisfy System (3) [17–19].

**Sensitivity analysis**

To quantify uncertainty in the underestimation factor, expansion factor, number of symptomatic influenza illnesses, and basic reproduction number, we utilize variability in parameter estimates in Table 3. Note that model parameter estimates appear as point values with the

exception of the recovery rate  $\gamma$  (Table 3). We solve system (4) according to each case definition (Modified-CDC and RI), country (US and Canada) and study year (2011-2012 and 2012-2013) with the mean value of  $\gamma$ , lower 95% confidence bound and upper 95% confidence bound on  $\gamma$ . The estimated value of  $p$  and  $\beta$  corresponding to the mean value of  $\gamma$  represent baseline results. The ranges of  $p$  and  $\beta$  obtained using the 95% confidence bounds of  $\gamma$  represent their sensitivity to an estimated 95% CI of the infectious period of influenza. We obtain baseline estimates of the expansion factor, number of symptomatic influenza illnesses and the basic reproduction number using baseline  $p$  and  $\beta$  estimates and methods in “[Expansion factor and symptomatic illnesses](#)” section. To retrieve an interval based on variation of recovery rate  $\gamma$ , we propagate the variability in  $p$  and  $\beta$  to each of the epidemiological parameters using methods in “[Expansion factor and symptomatic illnesses](#)” section.

**Sensitivity to underestimation in the RCT**

We have assumed perfect reporting and ascertainment of influenza virus infection within the RCT. Participants were instructed to contact their study site if they had any respiratory symptoms [17]. In addition, participants were contacted by a call center twice weekly (between the beginning of January and the end of February) or weekly until the end of illness surveillance (April 30 each year) [17]. In light of these frequent participant follow-ups in the RCT, we expect this assumption to hold. However, we consider the possibility of underestimation occurring within the RCT to be exhaustive in our analysis [17]. For details regarding the sensitivity of  $p$  and  $\beta$  to underestimation in the RCT, see Appendix C.

**Results**

**Disease transmission and burden estimates**

The estimates for Canada are displayed in Table 4. Using the laboratory-confirmed influenza counts associated with modified CDC-defined ILI within the RCT, we quantified the underestimation factor  $p = 2.6\%$  (2.1 – 3.2%) and  $p = 1.2\%$  (0.98 – 1.2%) in the 2011-2012 and 2012-2013 influenza seasons, respectively. These underestimation factors correspond to expansion factors of  $E = 37.9$  (31 – 46.8) and  $E = 84.8$  (69.1 – 102.6) in 2011-2012 and 2012-2013 influenza seasons, respectively (Fig. 2a). Equipped with expansion factors in each year, we estimated symptomatic influenza illnesses to be 390,000 (320,000 – 480,000) and 2.3 million (1.8 – 2.7 million) in 2011-2012 and 2012-2013 influenza seasons, respectively. Also in Canada, using the laboratory-confirmed influenza counts associated with RI within the RCT, we quantified the underestimation factor  $p = 0.51\%$  (0.42 – 0.63%) and  $p = 0.4\%$  (0.33 – 0.49%) in the 2011-2012 and 2012-2013 influenza seasons, respectively. These under-

**Table 4** Summary of estimates in Canada during 2011-2012 and 2012-2013 influenza seasons. Values reported as estimated baseline value and range from variation of recovery rate  $\gamma$  (see section “Sensitivity analysis” for the details of sensitivity analysis)

Seasonal influenza in Canada: 2011-2012 and 2012-2013 influenza seasons		
Laboratory-confirmed infections*	10,271	26,671
Estimates generated from modified CDC-defined influenza in RCT		
Season	2011-12	2012-13
Percent of influenza cases captured in surveillance	2.6% (2.1 – 3.2%)	1.2% (0.98 – 1.2%)
Expansion factor	37.9 (31.0 – 46.8)	84.8 (69.1 – 102.6)
Estimated symptomatic influenza illnesses	$3.9 \times 10^5$ ( $3.2 \times 10^5$ – $4.8 \times 10^5$ )	$2.3 \times 10^6$ ( $1.8 \times 10^6$ – $2.7 \times 10^6$ )
Basic reproduction number	1.22 (1.22 – 1.22)	1.19 (1.19 – 1.19)
Estimates generated from respiratory illness associated influenza in RCT		
Season	2011-12	2012-13
Percent of influenza cases captured in surveillance	0.51% (0.42 – 0.63%)	0.4% (0.33 – 0.49%)
Expansion factor	195.1 (159.3 – 236.1)	252.9 (206.3 – 306.1)
Estimated symptomatic influenza illnesses	$2.0 \times 10^6$ ( $1.6 \times 10^6$ – $2.4 \times 10^6$ )	$6.7 \times 10^6$ ( $5.5 \times 10^6$ – $8.2 \times 10^6$ )
Basic reproduction number	1.22 (1.22 – 1.22)	1.19 (1.19 – 1.19)

\*indicates infections recorded in Public Health Agency of Canada’s final seasonal FluWatch report [19]

estimation factors correspond to expansion factors of  $E = 200.1$  (163.4 – 243.9) and  $E = 376.1$  (306.7 – 455.1) in 2011-2012 and 2012-2013 influenza seasons, respectively. Equipped with expansion factors in each year, we estimated symptomatic influenza illnesses to be 2 million (1.6 – 2.4 million) and 6.7 million (5.5 – 8.2 million) in 2011-2012 and 2012-2013 influenza seasons, respectively.

The estimates for the US are displayed in Table 5. Using the laboratory-confirmed influenza associated with modified CDC-defined ILI within the RCT, we estimated  $p = 0.5\%$  (0.41 – 0.61%) and  $p = 0.27\%$  (0.2 – 0.33%) in the 2011-2012 and 2012-2013 seasons, respectively. These underestimation factors correspond to expansion factors of  $E = 200.1$  (163.4–243.9) and  $E = 376.1$  (306.7–455.1) in 2011-2012 and 2012-2013 influenza seasons, respectively (Fig. 2b). Equipped with expansion factors in each year, we estimated symptomatic influenza illnesses to be 5.4 million (4.4 – 6.9 million) and 29 million (23 – 34 million) in 2011-2012 and 2012-2013 influenza seasons, respectively. Also in the US, using the laboratory-confirmed influenza counts associated with RI within the RCT, we estimated  $p = 0.1\%$  (0.08 – 0.12%) and  $p = 0.09\%$  (0.07 – 0.11%) in the 2011-2012 and 2012-2013 seasons, respectively. These underestimation factors correspond to expansion factors of  $E = 1030.1$  (840.5 – 1247) and  $E = 1111.4$  (906.6 – 1345.3) in 2011-2012 and 2012-2013 influenza seasons, respectively. Equipped with expansion factors in each year, we estimated symptomatic influenza illnesses to be 28 million (23 – 34 million) and 84 million (69 – 102 million) in 2011-2012 and 2012-2013 influenza seasons, respectively.

We estimated the disease transmission rate  $\beta$  range to between 0.25 and 0.39 in each season and country, cor-

responding to basic reproduction numbers  $R_0$  ranging between 1.19 and 1.22 (Tables 4 and 5, RI case definition).

#### Sensitivity to underestimation in RCT

Recall we have assumed perfect reporting and ascertainment of influenza virus infection among the RCT participants and proposed to revisit this assumption with a sensitivity analysis. We have conducted a sensitivity analysis and present the details in Appendix C. Overall, the sensitivity analysis suggests that the results in “Results” section are robust to underestimation within the RCT [17]. We find that our estimates for  $\beta$  are weakly dependent on underestimation within the RCT [17]. Further, the captured fraction of influenza infections by virologic surveillance  $p$  is also robust to underestimation within the RCT [17], maintaining an order of magnitude while varying RCT underestimation over its full range. See Appendix C for a full presentation of this analysis.

#### Discussion

This study develops and illustrates a method utilizing a parameter estimation technique based on a mechanistic model and data synthesis to quantify the underestimation factor associated with season influenza and the number of symptomatic illnesses. These estimates take into account mechanistic detail of influenza transmission, vaccine effectiveness, relative vaccine efficacy of SD to HD from the RCT, and vaccine coverage. While this method does utilize RCT outcomes, the remaining data required to generate the estimates become publicly available by the end of the influenza season. This method can be utilized for epidemiological parameter estimation for infectious diseases, provided there is an appropriate form of active



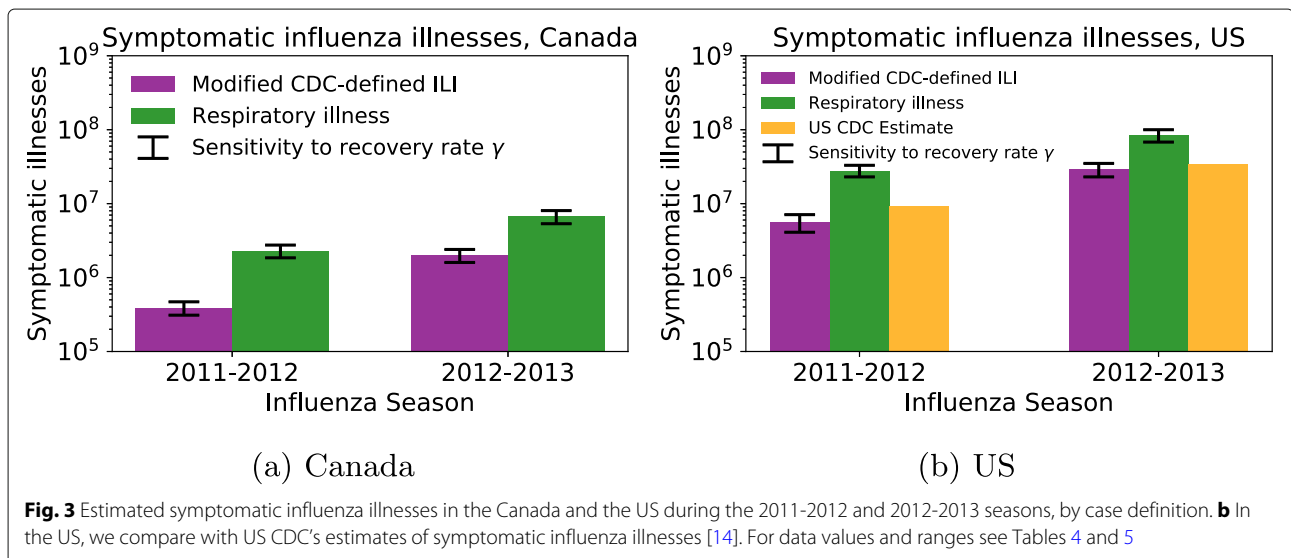
surveillance (e.g., clinical trial) data available. A coupled system of differential equations for the actively monitored population and general community can be developed to integrate surveillance epidemiological data to quantify key population-level parameters, including the underestimation factor.

Our estimates generated from the laboratory-confirmed influenza associated with modified CDC-defined ILI case definition are representative of the underestimation of cases which could be captured by the surveillance system (Tables 4 and 5, modified CDC-defined ILI case definition). In this light, our analysis indicates that the surveillance system in Canada captured 2.6% (2.1 – 3.2%) and 1.2% (0.98 – 1.2%) of symptomatic cases closely associated with modified CDC-defined ILI in 2011-2012 and 2012-2013 influenza seasons, respectively. In the US, the virologic surveillance system was estimated to capture 0.1% (0.08 – 0.12%) and 0.09% (0.07 – 0.11%) of symptomatic influenza cases closely associated with modified CDC-defined ILI in 2011-2012 and 2012-2013 influenza seasons, respectively. In each country, the percentage of captured cases decreases with an increase in laboratory-confirmed cases from passive surveillance, which may be due to laboratory testing capacity or changing testing practices (Tables 4 and 5, modified CDC-defined ILI).

The underestimation factors generated from laboratory-confirmed influenza associated with RI, i.e., a broader range of symptoms, provide an underestimation factor more closely representing the true symptomatic influenza underestimation factor and are apt to assess the true number of influenza illnesses (Tables 4 and 5, respiratory illness case definition). When considering a range of symptomatic influenza illness from the estimates using data specified by these two case definitions, we provide symptomatic influenza illness estimates which are

consistent with US CDC's in 2011-2012 and 2012-2013 (Fig. 3b). The US CDC estimate in 2012-2013 is near the low end of the range we estimated, which may be due to difference in methodology. In particular, this difference may be due to the nonlinear (exponential) relationship between the underestimation factor and laboratory-confirmed illnesses in surveillance data. An explanation to support this nonlinear relationship and our results could be that the increased number of influenza illnesses results in limited availability of laboratory tests, therefore a fewer proportion of cases were captured in virological surveillance.

We offer two sources of comparison from studies of the 2009 influenza pandemic. In an expert opinion from the 2009 influenza pandemic, the number of expected influenza infections per laboratory-confirmed infection ranged from 10 to 500 among experts [37]. For Canada, our estimates of the expansion factor lie in this range of 10 to 500 (Table 4). For the US, we estimate expansion factors greater than 500 using respiratory illness case definition (Table 5). Also, estimates of underestimation factors were between 0.1% and 0.7% during the 2009 influenza pandemic in Mexico [16]. Our estimates for the true underestimation in Canada and the US lie in this range (Table 4 and 5, respiratory illness case definition). In both cases, investigations of the 2009 pandemic influenza in Canada and Mexico align well with Canada; however, according to our analysis the underestimation in laboratory-confirmed surveillance in the US is more substantial. This is likely due to differences in surveillance systems, laboratory-testing protocols, or health-seeking behavior. We also note that epidemiological characteristics differ from pandemics to seasonal influenza and the level of awareness may have impacted health seeking behavior and diagnosis practices.



**Table 5** Summary of estimates in the US during 2011–2012 and 2012–2013 influenza seasons. Values reported as estimated baseline value and range from variation of recovery rate  $\gamma$  (see section “Sensitivity analysis” for the details of sensitivity analysis)

Seasonal influenza in the US: 2011–2012 and 2012–2013 influenza seasons		
Laboratory-confirmed infections*	27,012	75,772
Estimates generated from Modified CDC-defined ILI in RCT		
Season	2011–12	2012–13
Percent of influenza cases captured in surveillance	0.5% (0.41 – 0.61%)	0.27% (0.2 – 0.33%)
Expansion factor	200.1 (163.4 – 243.9)	376.1 (306.7 – 455.1)
Estimated symptomatic influenza illnesses	$5.4 \times 10^6$ ( $4.4 \times 10^6$ – $6.9 \times 10^6$ )	$2.9 \times 10^7$ ( $2.3 \times 10^7$ – $3.4 \times 10^7$ )
Basic reproduction number	1.19 (1.19 – 1.19)	1.16 (1.16 – 1.16)
Estimates generated from respiratory illness associated influenza in RCT		
Season	2011–12	2012–13
Percent of influenza cases captured in surveillance	0.1% (0.08 – 0.12%)	0.09% (0.07 – 0.11%)
Expansion factor	1030.1 (840.5 – 1247.0)	1111.4 (906.6 – 1345.3)
Estimated symptomatic influenza illnesses	$2.8 \times 10^7$ ( $2.3 \times 10^7$ – $3.4 \times 10^7$ )	$8.4 \times 10^7$ ( $6.9 \times 10^7$ – $1.02 \times 10^8$ )
Basic reproduction number	1.20 (1.20 – 1.20)	1.19 (1.19 – 1.19)
US CDC estimated symptomatic influenza illnesses	$9.3 \times 10^6$	$3.4 \times 10^7$

\*Indicates infections recorded by the US Center for Disease and Control’s FluView reports [18]

The basic reproduction number ranged from 1.19 - 1.22 for estimates associated with RI case definition, which more closely represent the true  $R_0$ . While these basic reproduction numbers align tightly in the US and Canada, our estimates for  $R_0$  are higher in Canada in all influenza seasons (Tables 4 and 5). A recent systematic review found an interquartile range of reproduction numbers for seasonal influenza of 1.19–1.37, with median value 1.27 [38]. Our estimates for  $R_0$  are in line with typical findings representative of seasonal influenza [38].

We estimate the symptomatic influenza illnesses in the US and Canada; however, the number of asymptomatic influenza illnesses has not been assessed in this work. A recent review of estimates for the fraction of asymptomatic infections vary widely; however, values from 20%–50% are typical [39]. These estimates indicate that the number of asymptomatic influenza illnesses may be substantial. Even so, based on our scan of the literature, there is no clear consensus on the contribution of asymptomatic individuals to influenza transmission. Future studies may attempt to estimate the total number influenza illnesses, symptomatic and asymptomatic, to provide a more accurate assessment of influenza illnesses; influenza underestimation; and the force of infection. As a result of this simplification, we may underestimate  $R_0$  and hence the force of infection.

There are several limitations in the current study. There are differences in symptomatic attack rates, disease presentation, and health-seeking behaviour between age groups. As a result, the symptomatic reporting fraction likely varies from age group to age group. Future study may be extended to address these age group disparities by quantifying the underestimation and expansion factors

for each age group. Another limitation to this study is the assumed homogeneous mixing between age groups; however, this may also be addressed in a future study. In fact, the RCT data may allow the contact preference between age groups (i.e. the contact rate between age groups 0–64 and 65+, and vice-versa) to be estimated. In this light, the methods presented herein may be used as a method to validate empirical social contact data obtained from surveys or existing contact mixing patterns. In addition to age-specific heterogeneity, there may be spatial heterogeneity in contact mixing, surveillance systems, and social behavior which impacts case ascertainment and reporting rates. In this light, it may be advantageous to conduct analysis at more granular scales (e.g. city, state, or provincial levels) to more accurately capture transmission and underestimation at these sub-national levels. The RCT was conducted in the US and Canada, and we use aggregate case counts and RCT participant counts from US and Canada. In other words, the RCT is assumed to take place in the US or Canada. The underestimation factor could also be separated into several multipliers, as in prior works [13]. An improvement could be made to account for laboratory test accuracy; laboratory-confirmed cases from the RCT and surveillance reports can be preprocessed for the method to generate more representative estimates. Overall, this work lays a functional framework to expand as the research questions at hand require.

## Conclusions

We develop a method for quantifying the underestimation factor and disease transmission rate by integrating several data sources including RCT outcomes. We use our method to assess surveillance system capacity, number of

symptomatic influenza illnesses, and  $R_0$  in the US and Canada during 2011-2012 and 2012-2013 seasons. The utilization of outcomes from RCTs (in this case a comparative vaccine RCT) may allow for the extraction of additional information characterizing an epidemic which may not be possible by limiting data usage to national influenza surveillance reports. This work illustrates the point that monitoring a representative group within a population may aid in effectively modelling the transmission of infectious diseases such as influenza. In general, the utilization of available surveillance data may increase the capabilities of disease models and broaden their power to draw inferences. A formal structural and practical identifiability analysis should be carried out to rigorously address these details of parameter identification.

It is key to provide more accurate methods to estimate the annual incidence of influenza to guide evidence-based immunization policy-making. In this light, it is vital to develop more accurate mathematical models of influenza transmission, and to accurately evaluate the impact of influenza vaccination. These methods may contribute to and enable the design of optimal vaccination programs that best reduce the annual incidence of seasonal influenza as well as associated hospitalizations, medical visits and deaths.

**Appendix A: Mathematical details for estimating  $p$  and  $\beta$ .**

Here we present the details for estimating  $p$  and  $\beta$ . We also introduce the following notation for convenience and readability; let a hat ( $\hat{\cdot}$ ) over a variable denote integration over all nonnegative time. For example,  $\hat{I} = \int_0^\infty I(s)ds$ .

Integrating the first equation of submodel (1) we have

$$\ln \frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} = -\epsilon_{+,SD}\beta(\hat{I}_{+,SD}^C + \hat{I}_{+,HD}^C + \hat{J}). \tag{7}$$

Similarly, taking the second equation from submodel (1), we have:

$$\ln \frac{V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)} = -\epsilon_{+,HD}\beta(\hat{I}_{+,SD}^C + \hat{I}_{+,HD}^C + \hat{J}). \tag{8}$$

For finding  $\hat{I}_{+,SD}^C, \hat{I}_{+,HD}^C$ , adding the first and third equations, & second and fourth of submodel (1) followed by integrating both sides give us:

$$\hat{I}_{+,SD}^C = -\frac{V_{+,SD}^C(\infty) - V_{+,SD}^C(0)}{\gamma},$$

$$\hat{I}_{+,HD}^C = -\frac{V_{+,HD}^C(\infty) - V_{+,HD}^C(0)}{\gamma},$$

provided that  $I_{+,SD}^C(0) = I_{+,SD}^C(\infty) = I_{+,HD}^C(0) = I_{+,HD}^C(\infty) = 0$ . This, along with  $pJ = R$  (where  $J = I_-^O + I_+^O$ ), from Eqs. 7 and (8) we have:

$$\frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} = e^{\frac{\epsilon_{+,SD}\beta}{\gamma}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R})}$$

and

$$\frac{V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)} = e^{\frac{\epsilon_{+,HD}\beta}{\gamma}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R})}.$$

Thus, we define the following functions:

$$f(\beta, p) = V_{+,SD}^C(\infty) - V_{+,SD}^C(0)e^{\frac{\epsilon_{+,SD}\beta}{\gamma}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R})} \tag{9}$$

$$h(\beta, p) = V_{+,HD}^C(\infty) - V_{+,HD}^C(0)e^{\frac{\epsilon_{+,HD}\beta}{\gamma}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R})} \tag{10}$$

Now, from model (2) we have the following five equations:

$$\frac{\dot{S}_-^O}{S_-^O} = -\beta J, \quad \frac{\dot{S}_+^O}{S_+^O} = -\beta J, \quad \frac{\dot{V}_{-,SD}^O}{V_{-,SD}^O} = -\epsilon_{-,SD}\beta J,$$

$$\frac{\dot{V}_{+,SD}^O}{V_{+,SD}^O} = -\epsilon_{+,SD}\beta J, \quad \frac{\dot{V}_{+,HD}^O}{V_{+,HD}^O} = -\epsilon_{+,HD}\beta J.$$

Recall  $pJ = R$ , where  $p$  is the underestimation factor. Integrating the equations above yields the respective final size relations:

$$S_-^O(\infty) = S_-^O(0)e^{-(\beta p^{-1}\hat{R})},$$

$$S_+^O(\infty) = S_+^O(0)e^{-(\beta p^{-1}\hat{R})},$$

$$V_{-,SD}^O(\infty) = V_{-,SD}^O(0)e^{-(\beta\epsilon_{-,SD}p^{-1}\hat{R})},$$

$$V_{+,SD}^O(\infty) = V_{+,SD}^O(0)e^{-(\beta\epsilon_{+,SD}p^{-1}\hat{R})},$$

$$V_{+,HD}^O(\infty) = V_{+,HD}^O(0)e^{-(\beta\epsilon_{+,HD}p^{-1}\hat{R})}.$$

Further, by summing all the equations of model (2), we obtain the following equality:

$$\dot{S}_-^O + \dot{S}_+^O + \dot{V}_{-,SD}^O + \dot{I}_+^O + \dot{V}_{+,SD}^O + \dot{V}_{+,HD}^O + \dot{I}_-^O = -\gamma J.$$

Integrating both sides of the above equations with the assumption that there are no infections at  $t_0$  and  $t_\infty$  gives:

$$S_-^O(\infty) - S_-^O(0) + S_+^O(\infty) - S_+^O(0) + (V_{-,SD}^O(\infty) - V_{-,SD}^O(0)) + (V_{+,SD}^O(\infty) - V_{+,SD}^O(0)) + (V_{+,HD}^O(\infty) - V_{+,HD}^O(0)) = -\gamma p^{-1}\hat{R}. \tag{11}$$

Substituting values of final sizes from above relations and  $p\hat{J} = \hat{R}$  into (11):

$$\begin{aligned} & S_-^O(0)e^{-(\beta p^{-1}\hat{R})} - S_-^O(0) + \left( V_{-,SD}^O(0)e^{-(\beta\epsilon_{-,SD}p^{-1}\hat{R})} \right. \\ & \left. - V_{-,SD}^O(0) \right) + \left( V_{+,SD}^O(0)e^{-(\beta\epsilon_{+,SD}p^{-1}\hat{R})} - V_{+,SD}^O(0) \right) \\ & + \left( V_{+,HD}^O(0)e^{-(\beta\epsilon_{+,HD}p^{-1}\hat{R})} - V_{+,HD}^O(0) \right) = -\gamma p^{-1}\hat{R} \end{aligned}$$

Thus we define:

$$\begin{aligned} g(\beta, p) = & e^{-(\beta p^{-1}\hat{R})} \left( S_-^O(0) + S_+^O(0) \right) + V_{-,SD}^O(0)e^{-(\beta\epsilon_{-,SD}p^{-1}\hat{R})} \\ & + V_{+,SD}^O(0) \left( e^{-(\beta\epsilon_{+,SD}p^{-1}\hat{R})} \right) + V_{+,HD}^O(0) \left( e^{-(\beta\epsilon_{+,HD}p^{-1}\hat{R})} \right) \\ & - \left( S_-^O(0) + S_+^O(0) + V_{-,SD}^O(0) + V_{+,SD}^O(0) + V_{+,HD}^O(0) \right) + \gamma p^{-1}\hat{R}. \end{aligned} \tag{12}$$

With this definition, we obtain the following system of equations, which  $\beta$  and  $p$  must satisfy:

$$\begin{aligned} f(\beta, p) &= 0, \\ g(\beta, p) &= 0. \end{aligned} \tag{13}$$

Suppose that  $\beta$  and  $p$  satisfy  $f(\beta, p) = 0$ , then we have:

$$\beta = -\ln \frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} \cdot \frac{\gamma}{\epsilon_{+,SD}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0)) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R}}$$

and that  $\beta$  and  $p$  also satisfy  $h(\beta, p) = 0$ , that is

$$\beta = -\ln \frac{V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)} \cdot \frac{\gamma}{\epsilon_{+,HD}(V_{+,SD}^C(\infty) - V_{+,SD}^C(0)) + V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{R}}.$$

Note that for both of these equalities to hold, it is required that

$$\frac{1}{\epsilon_{+,SD}} \ln \frac{V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} = \frac{1}{\epsilon_{+,HD}} \ln \frac{V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)}. \tag{14}$$

Equation (10) does hold since this is precisely the expression used to estimate  $\epsilon_{+,HD}$  from vaccine RCT information and  $\epsilon_{+,SD}$ .

Since all known variables in  $f$  and  $g$  have been estimated except for  $\beta$  and  $p$ , we can obtain estimates for  $\beta$  and  $p$  by solving (15) (e.g., numerically).

$$\begin{aligned} f(\beta, p) &= 0, \\ g(\beta, p) &= 0. \end{aligned} \tag{15}$$

Solving System (15) yields estimates for  $p$  and  $\beta$ , which we provide with numerical methods in Section “[Disease transmission and burden estimates](#)”. Lastly, we note that Equations (9) and (12) remain intact if a latent compartment were to be added to models (1) and (2) to account for the incubation period of influenza.

## Appendix B: Detailed parameter estimation

Here we present the details for the calculations of parameter values  $\epsilon_{+,SD}$ ,  $\epsilon_{+,HD}$  and  $\epsilon_{-,SD}$  for influenza seasons 2011-2012 and 2012-2013. For this purpose, we utilize an approximation which relates vaccine-modified susceptibility to vaccine effectiveness (VE) [36]. We also leverage influenza vaccine effectiveness (VE) from several studies in the US and Canada.

### 2011-2012 Influenza season

#### Canada

*Informing  $\epsilon_{-,SD}$ :* First, we inform  $\epsilon_{+,SD}$  using VE estimates. In Canada, the VE against Real-Time Polymerase Chain Reaction (Real-time PCR) confirmed influenza in a test-negative case-control study was estimated to be 55% among all participants of all ages [25]. We use the approximation  $\epsilon = 1 - VE$  to estimate  $\epsilon_{-,SD}$ ; that is  $\epsilon_{-,SD} = 1 - 0.55 = 0.45$  [36].

*Informing  $\epsilon_{+,SD}$ :* In the same test-negative case-control study, the VE against Real-time PCR confirmed influenza among participants aged 50+ was estimated to be 58% [25]. We use this figure to estimate  $\epsilon_{+,SD}$  using the approximation  $\epsilon = 1 - VE$  to find  $\epsilon_{+,SD} = 1 - 0.58 = 0.42$  [36].

#### US

*Informing  $\epsilon_{-,SD}$ :* In the US, the VE against medically-attended influenza among all ages was estimated to be 47% [26]. We now use the approximation  $\epsilon = 1 - VE$  to estimate  $\epsilon_{-,SD}$  as follows:  $\epsilon_{-,SD} = 1 - 0.47 = 0.53$ .

*Informing  $\epsilon_{+,SD}$ :* The VE against medically-attended influenza among those aged 65+ was estimated to be 43% in the US in 2011-2012 [26]. With the relationship  $\epsilon_{+,SD} = 1 - VE$  we calculate  $\epsilon_{-,SD} = 0.57$  [36].

*Informing  $\epsilon_{+,HD}$ :* Lastly, to estimate the high-dose vaccine-modified susceptibility we use Equation (3) in the main text informed with RCT results to calculate  $\epsilon_{+,HD}/\epsilon_{+,SD}$  [17].

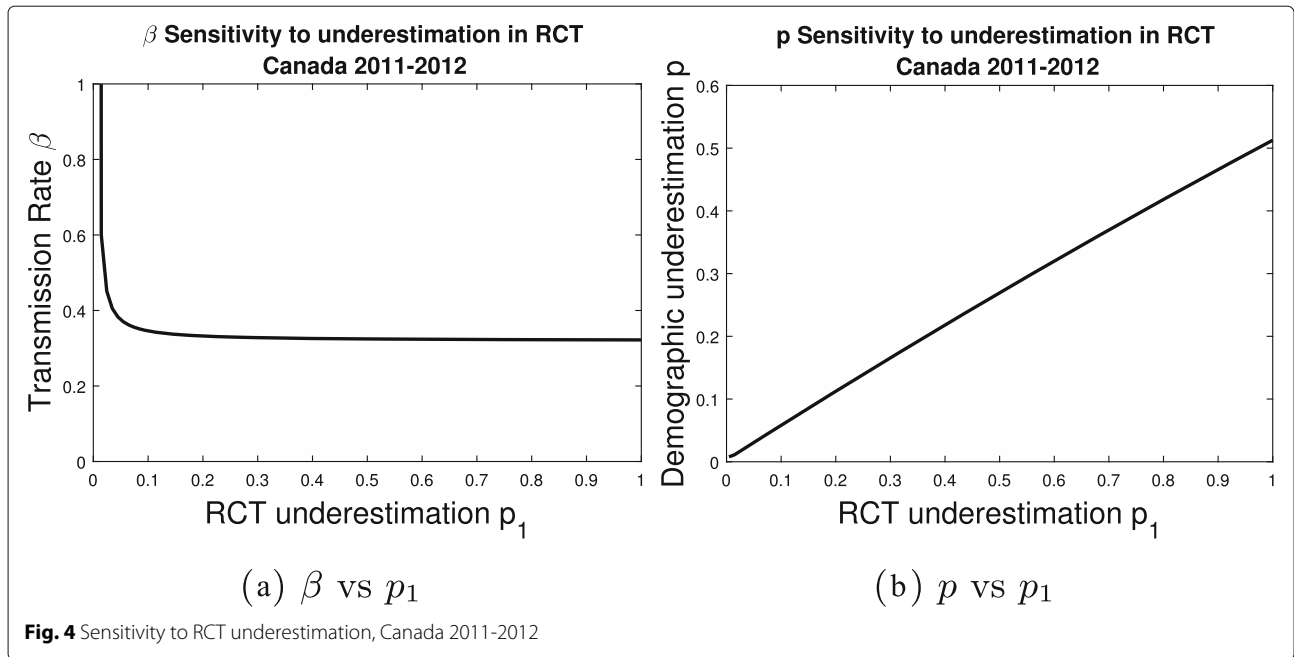
$$\frac{\epsilon_{+,HD}}{\epsilon_{+,SD}} = \frac{\log(7230/7253)}{\log(7202/7244)} = 0.55.$$

Note that with the ratio of two vaccine-modified susceptibilities and  $\epsilon_{+,SD}$  known, we can calculate  $\epsilon_{+,HD}$ . For the US, we then have  $\epsilon_{+,HD} = 0.55 \times 0.53 = 0.29$ .

### 2012-2013

#### Canada

*Informing  $\epsilon_{-,SD}$ :* We inform  $\epsilon_{-,SD}$  using the VE against PCR-confirmed influenza from a test-negative case control study in Canada [24]. In this study, the VE against PCR-confirmed influenza was estimated to be 51% [24]. Now, we use the approximation  $\epsilon = 1 - VE = 1 - 0.51 = 0.49$  [24].



*Informing*  $\epsilon_{+,SD}$ : In the same test-negative case control study, VE was estimated to be 47% among participants aged 50+. As a result, we estimate  $\epsilon_{+,SD}$  using the approximation  $\epsilon = 1 - VE$  to find  $\epsilon_{+,SD} = 1 - 0.47 = 0.53$  [36].

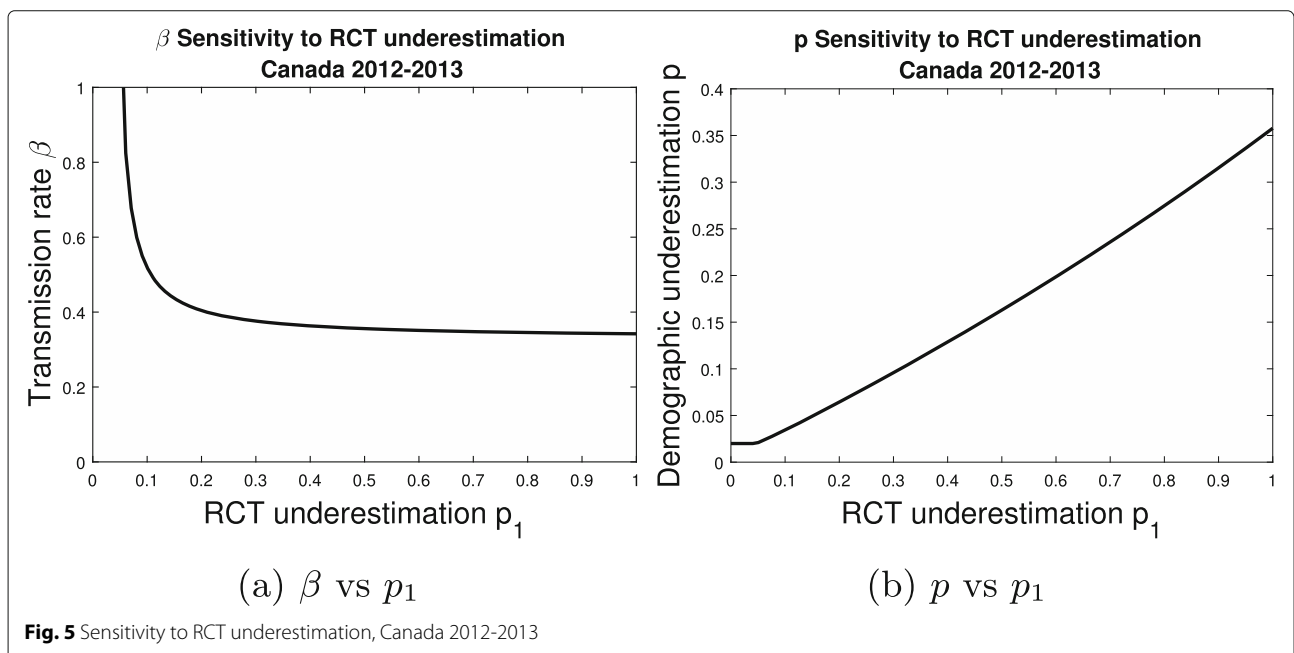
**US**

*Informing*  $\epsilon_{-,SD}$ : First, we inform  $\epsilon_{-,SD}$  using the estimates for the VE against medically-attended influenza over all ages in the US of 49% [27]. We then calculate  $\epsilon_{-,SD} = 1 -$

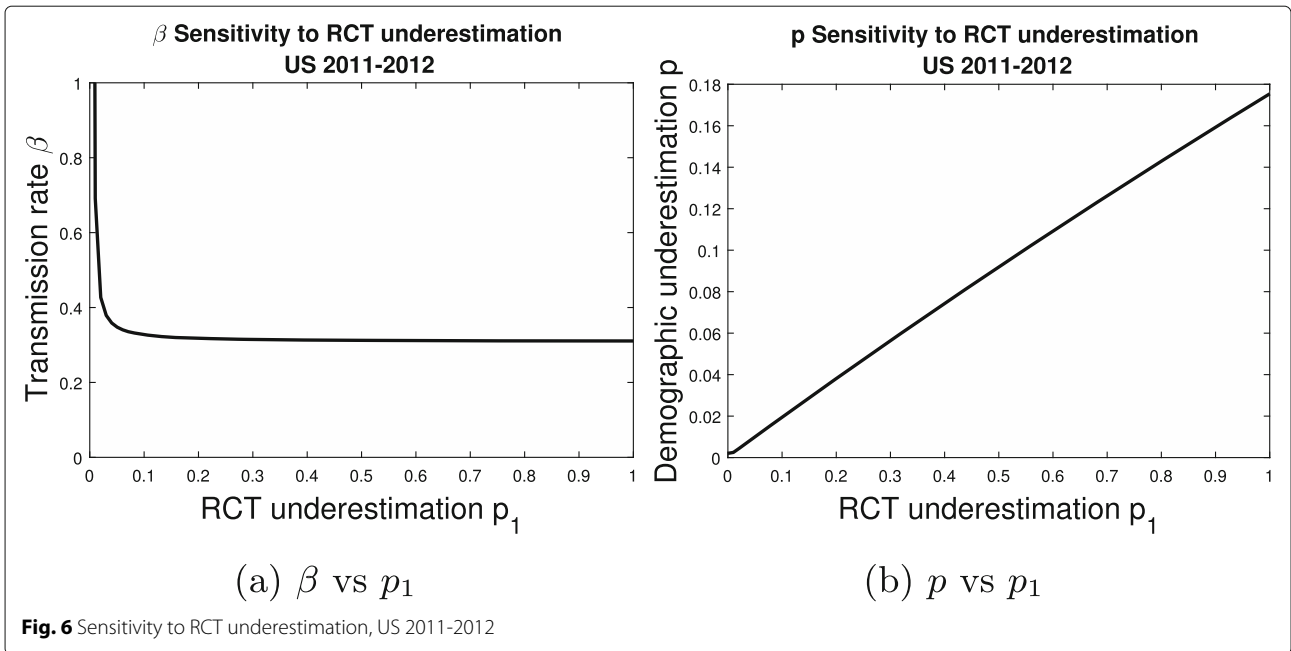
$0.49 = 0.51$  using the approximation  $\epsilon_{+,SD} = 1 - VE$  [36].

*Informing*  $\epsilon_{+,SD}$ : In the same US study, VE against medically-attended influenza was estimated to be 29% among those aged 65+ [27]. We then calculate  $\epsilon_{+,SD} = 1 - 0.29 = 0.71$  using the approximation  $\epsilon_{+,SD} = 1 - VE$  [36].

*Informing*  $\epsilon_{+,HD}$ : Now, to estimate the high-dose vaccine-modified susceptibility we first use Equation (3) in the main text informed by RCT results to calculate  $\epsilon_{+,SD}/\epsilon_{+,HD}$  [17].







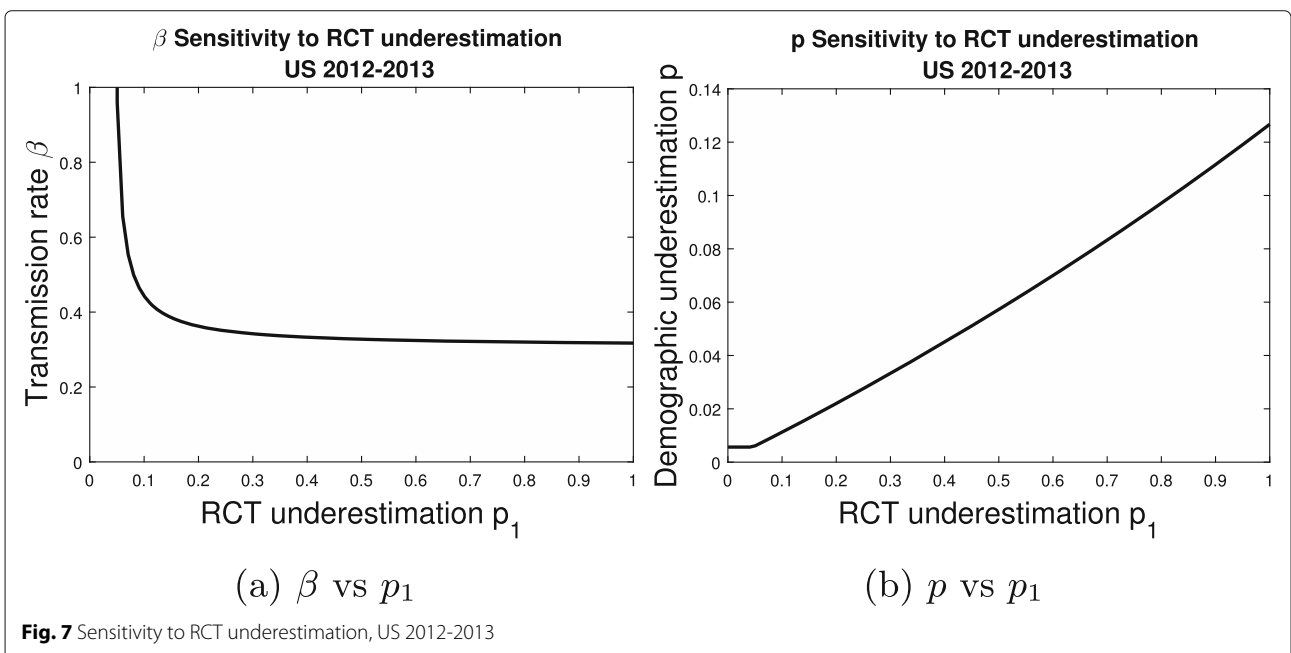
$$\frac{\epsilon_{+,HD}}{\epsilon_{+,SD}} = \frac{\log(8532/8737)}{\log(8490/8749)} = 0.79.$$

We use the known ratio:  $\epsilon_{+,SD}/\epsilon_{+,HD}$  to calculate  $\epsilon_{+,HD} = 0.79 * 0.71 = 0.56$ . In the US we have  $\epsilon_{+,HD} = 0.56$ .

**Appendix C: Accounting for underestimation in the RCT**

In this section, we present the details of the sensitivity analysis exploring the impact of underestimation in the

RCT on study results. Specifically, we investigate the sensitivity of  $p$  and  $\beta$  on underestimation in the RCT [17]. To consider underestimation in the RCT, we introduce a underestimation factor  $p_1$  in the RCT. As before,  $p$  remains the underestimation factor in the general community. We also assume that RCT participants provided with standard-dose and high-dose vaccines are equally underestimated. Lastly, we define  $\delta$  to be the fraction of true remaining vaccinated participants at the end of the RCT.



**Method**

The total number of observed infections in the RCT can be written as:

$$\bar{I} := V_{+,SD}^C(0) + V_{+,HD}^C(0) - V_{+,SD}^C(\infty) - V_{+,HD}^C(\infty). \quad (16)$$

The above equation can also be formulated in terms of  $p_1$  and  $\delta$ , that is:

$$\bar{I} = p_1 [V_{+,SD}^C(0) + V_{+,HD}^C(0) - \delta(V_{+,SD}^C(\infty) + V_{+,HD}^C(\infty))]. \quad (17)$$

The relationship between  $p_1$  and  $\delta$  can be determined by equating (16) and (17) which yields:

$$\delta = \frac{1}{V_{+,SD}^C(\infty) + V_{+,HD}^C(\infty)} [V_{+,SD}^C(0) + V_{+,HD}^C(0) - p_1^{-1} (V_{+,SD}^C(0) + V_{+,HD}^C(0) - V_{+,SD}^C(\infty) - V_{+,HD}^C(\infty))]. \quad (18)$$

Under these conditions, by conducting the same analysis of model (1) as shown in Appendix A, we derive the condition:

$$\frac{1}{\epsilon_{+,SD}} \ln \frac{\delta V_{+,SD}^C(\infty)}{V_{+,SD}^C(0)} = \frac{1}{\epsilon_{+,HD}} \ln \frac{\delta V_{+,HD}^C(\infty)}{V_{+,HD}^C(0)}. \quad (19)$$

Continuing to follow the analysis from Appendix A, we arrive at the following system of equations:

$$\begin{aligned} \bar{f}(\beta, p, \delta) &= \delta V_{+,SD}^C(\infty) - V_{+,SD}^C(0) e^{\frac{\epsilon_{+,SD}\beta}{\gamma} (\delta V_{+,SD}^C(\infty) - V_{+,SD}^C(0) + \delta V_{+,HD}^C(\infty) - V_{+,HD}^C(0) - p^{-1}\gamma\hat{r})}, \\ \bar{g}(\beta, p) &= e^{-(\beta p^{-1}\hat{r})} (S_{+,SD}^O(0) + S_{+,SD}^O(0)) + V_{+,SD}^O(0) e^{-(\beta\epsilon_{-,SD}p^{-1}\hat{r})} \\ &\quad + V_{+,SD}^O(0) e^{-(\beta\epsilon_{+,SD}p^{-1}\hat{r})} + V_{+,HD}^O(0) e^{-(\beta\epsilon_{+,HD}p^{-1}\hat{r})} \\ &\quad - (S_{+,SD}^O(0) + S_{+,SD}^O(0) + V_{+,SD}^O(0) + V_{+,SD}^O(0) + V_{+,HD}^O(0)) + \gamma p^{-1}\hat{r}. \end{aligned} \quad (20)$$

We analyze the robustness of the estimates of  $\beta$  and  $p$  by varying  $\delta$  in the above system of equations. Exploring sensitivity with respect to  $\delta$  gives insight into how underestimation in the RCT affects our results in Section “Results”. Specifically, for a given  $p_1$ , we calculate a corresponding  $\delta$  then solve the above system numerically.

**Results**

The relationship between the RCT underestimation factor,  $p_1$ , with  $\beta$  and  $p$  estimates is shown in Figs. 4, 5, 6, and 7. We explore the sensitivity of  $p_1$  on  $p$  and  $\beta$  for each context of interest; both regionally in the US and Canada during 2011-12 and 2012-13 influenza seasons. We present the sensitivity analysis using the laboratory-confirmed influenza associated with RI case definition outcomes in the RCT.

Overall,  $p$  and  $\beta$  are robust to underestimation in the RCT. We see that the effect of underestimation in the RCT is negligible on the transmission rate  $\beta$  for reasonable values of  $p_1$ . On the other hand,  $p$  is also weakly dependent on  $p_1$ . By varying  $p_1$  over its entire range,  $p$  remains to be on the same order of magnitude.

**Abbreviations**

SVIR: Susceptible-vaccinated-infected-recovered; RCT: Randomized-controlled trial; PHAC: Public Health Agency of Canada; CDC: Centers for Disease Control and Prevention; SD: Standard-dose vaccine; HD: High-dose vaccine; Real-time PCR: Real-Time Polymerase Chain Reaction; RI: Respiratory illness

**Acknowledgements**

Special thanks to the Fields Institute and the Centre for Quantitative Analysis (CQAM) for hosting the Industrial Problem Solving Workshop for which this research originated.

**Authors’ contributions**

CC, MA, SA, AFT, and ZM prepared the manuscript. Research was conceptualized by SA, JL, JH, ZM, TS, and JW. Numerical simulations conducted by ZM, SA, and CC. All authors contributed significantly to this work. All authors read and approved the final manuscript.

**Funding**

ZM, MA, SA, JW, CC, JDK have received support from Sanofi Pasteur and the Natural Science and Engineering Council of Canada. JW is supported by the Natural Sciences and Engineering Research Council of Canada (Grant number:105588-2011) and the Canada Research Chairs Program (Grant number:230720). JH is supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grants Program and Discovery Accelerator Supplements (DAS) Program. JH is also supported by the York Research Chairs Program. This project is supported by the NSERC/Sanofi Industrial Research Chair Program in Vaccine Mathematics, Modelling and Manufacturing.

**Availability of data and materials**

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

AC, TS, JL, AFT, ET, LC, and BS are employees of Sanofi Pasteur which is a manufacturer of influenza vaccines.

**Author details**

<sup>1</sup>Department of Mathematics and Statistics, York University, M3J 1P3 Toronto, Canada. <sup>2</sup>Laboratory for Industrial and Applied Mathematics, York University, M3J 1P3 Toronto, ON, Canada. <sup>3</sup>Centre for Disease Modelling, York University, M3J 1P3 Toronto, ON, Canada. <sup>4</sup>Fields-CQAM Mathematics for Public Health Laboratory, York University, M3J 1P3 Toronto, ON, Canada. <sup>5</sup>University of Alberta, T6G 2R3 Edmonton, AB, Canada. <sup>6</sup>University of Guelph, N1G 2W1 Guelph, ON, Canada. <sup>7</sup>McMaster University, L8S 4L8 Hamilton, ON, Canada. <sup>8</sup>Lunenfeld-Tanenbaum Research Institute, Sinai Health System, M5G 1X5 Toronto, ON, Canada. <sup>9</sup>Sanofi Pasteur, M2R 3T4 Toronto, Canada. <sup>10</sup>Department of Molecular Genetics, M5S 1A8 Toronto, ON, Canada. <sup>11</sup>Sanofi Pasteur, Lyon, France. <sup>12</sup>University of Manitoba, Department of Mathematics, R3T 2N2 Winnipeg, MB, Canada. <sup>13</sup>Leslie Dan School of Pharmacy, University of Toronto, M5S 3M2 Toronto, ON, Canada. <sup>14</sup>Sanofi Pasteur, 18370 Swiftwater, PA, USA.

Received: 7 March 2020 Accepted: 28 May 2020

Published online: 10 July 2020

## References

- Gibbons CL, Mangen M-JJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, Peterson KL, Stuurman AL, Cassini A, Fèvre EM, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*. 2014;14(1):147.
- Poehling KA, Edwards KM, Weinberg GA, Szilagyi P, Staat MA, Iwane MK, Bridges CB, Grijalva CG, Zhu Y, Bernstein DI, et al. The underrecognized burden of influenza in young children. *N Eng J Med*. 2006;355:31–40.
- Grijalva C, Weinberg G, Bennett N, Staat M, Craig A, Dupont W, Iwane M, Postema A, Schaffner W, Edwards K, et al. Estimating the undetected burden of influenza hospitalizations in children. *Epidemiol Infect*. 2007;135(6):951–8.
- Barker WH, Mullooly JP. Underestimation of the role of pneumonia and influenza in causing excess mortality. *Am J of Public Health*. 1981;71(6):.
- Ortiz JR, Neuzil KM, Shay DK, Rue TC, Neradilek MB, Zhou H, Seymour CW, Hooper LG, Cheng P-Y, Goss CH, et al. The burden of influenza-associated critical illness hospitalizations. *Crit Care Med*. 2014;42(11):2325–32.
- Ortiz JR, Neuzil KM, Cooke CR, Neradilek MB, Goss CH, Shay DK. Influenza pneumonia surveillance among hospitalized adults may underestimate the burden of severe influenza disease. *PLoS one*. 2014;9(11):.
- Goldstein E, et al. Improving the estimation of influenza-related mortality over a seasonal baseline. *Epidemiology*. 2012;23(6):829–38.
- Matias G, Taylor R, Haguinet F, Schuck-Paim C, Lustig R, Shinde V. Estimates of mortality attributable to influenza and rsv in the united states during 1997–2009 by influenza type or subtype, age, cause of death, and risk status. *Influenza Other Respir Viruses*. 2014;8(5):507–15.
- Millman AJ, Reed C, Kirley PD, Aragon D, Meek J, Farley MM, Ryan P, Collins J, Lynfield R, Baumbach J, et al. Improving accuracy of influenza-associated hospitalization rate estimates. *BMC Infect Dis*. 2015;21(9):.
- Schanzer DL, Langley JM, Tam TW. Hospitalization attributable to influenza and other viral respiratory illnesses in canadian children. *Pediatr Infect Dis J*. 2006;25(9):795–800.
- Schanzer DL, Langley JM, Tam TW. Role of influenza and other respiratory viruses in admissions of adults to canadian hospitals. *Influenza Other Respir Viruses*. 2008;2(1):1–8.
- Schanzer DL, McGeer A, Morris K. Statistical estimates of respiratory admissions attributable to seasonal and pandemic influenza for canada. *Influenza Other Respir Viruses*. 2013;7(5):799–808.
- Reed C, Angulo FJ, Swerdlow DL, Lipsitch M, Meltzer MI, Jernigan D, Finelli L. Estimates of the prevalence of pandemic (h1n1) 2009, united states, april–july 2009. *Emerg Infect Dis*. 2009;15(12):2004.
- US Centers for Disease Control and Prevention. Disease Burden of Influenza. <https://www.cdc.gov/flu/about/burden/index.html>. Accessed 10 Sept 2019.
- McDonald SA, Presanis AM, De Angelis D, van der Hoek W, Hooiveld M, Donker G, Kretzschmar ME. An evidence synthesis approach to estimating the incidence of seasonal influenza in the netherlands. *Influenza Other Respir Viruses*. 2014;8(1):33–41.
- Chong K, Fong H, Zee C. Estimating the incidence reporting rates of new influenza pandemics at an early stage using travel data from the source country. *Epidemiol Infect*. 2014;142(5):955–63.
- DiazGranados CA, Dunning AJ, Kimmel M, Kirby D, Treanor J, Collins A, Pollak R, Christoff J, Earl J, Landolfi V, et al. Efficacy of high-dose versus standard-dose influenza vaccine in older adults. *N Engl J Med*. 2014;371(7):635–45.
- Centers for Disease Control and Prevention. Weekly U.S. Influenza Surveillance Report. <https://www.cdc.gov/flu/weekly/index.htm>. Accessed 8 May 2019.
- Public Health Agency of Canada. Weekly Influenza Reports. <https://www.canada.ca/en/public-health/services/diseases/flu-influenza/influenza-surveillance/weekly-influenza-reports.html>. Accessed 9 May 2019.
- Statistics Canada. Census Profile. <https://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/index.cfm?Lang=E>. Accessed 8 May 2019.
- Statistics Canada. Census Profile. <https://www12.statcan.gc.ca/census-recensement/2012/dp-pd/prof/index.cfm?Lang=E>. Accessed 8 May 2019.
- US Census Bureau. Age and Sex Composition in the United States: 2011. <https://www.census.gov/data/tables/2011/demo/age-and-sex/2011-age-sex-composition.html>. Accessed 7 May 2019.
- US Census Bureau. Age and Sex Composition in the United States: 2012. <https://www.census.gov/data/tables/2012/demo/age-and-sex/2012-age-sex-composition.html>. Accessed 7 May 2019.
- Skowronski DM, Janjua NZ, De Serres G, Sabaiduc S, Eshaghi A, Dickinson JA, Fonseca K, Winter A-L, Gubbay JB, Krajden M, et al. Low 2012–13 influenza vaccine effectiveness associated with mutation in the egg-adapted h3n2 vaccine strain not antigenic drift in circulating viruses. *PLoS one*. 2014;9(3):92153.
- Skowronski DM, Janjua NZ, Sabaiduc S, De Serres G, Winter A-L, Gubbay JB, Dickinson JA, Fonseca K, Charest H, Bastien N, et al. Influenza a/subtype and b/lineage effectiveness estimates for the 2011–2012 trivalent vaccine: cross-season and cross-lineage protection with unchanged vaccine. *J Infect Dis*. 2014;210(1):126–37.
- Ohmit SE, Thompson MG, Petrie JG, Thaker SN, Jackson ML, Belongia EA, Zimmerman RK, Gaglani M, Lamerato L, Spencer SM, et al. Influenza vaccine effectiveness in the 2011–2012 season: protection against each circulating virus and the effect of prior vaccination on estimates. *Clin Infect Dis*. 2013;58(3):319–27.
- McLean HQ, Thompson MG, Sundaram ME, Kieke BA, Gaglani M, Murthy K, Piedra PA, Zimmerman RK, Nowalk MP, Raviotta JM, et al. Influenza vaccine effectiveness in the united states during 2012–2013: variable protection by age and virus type. *J Infect Dis*. 2014;211(10):1529–40.
- Izurieta HS, Thadani N, Shay DK, Lu Y, Maurer A, Foppa IM, Franks R, Pratt D, Forshee RA, MaCurdy T, et al. Comparative effectiveness of high-dose versus standard dose influenza vaccines in us residents aged 65 years and older from 2012 to 2013 using medicare data: a retrospective cohort analysis. *Lancet Infect Dis*. 2015;15:293–300.
- National Advisory Committee on Immunization (NACI). Vaccine Coverage Amongst Adult Canadians: Results from the 2012 Adult National Immunization Coverage (aNIC) Survey. <https://www.canada.ca/en/public-health/services/immunization/vaccine-coverage-amongst-adult-canadians-results-2012-adult-national-immunization-coverage-anic-survey.html>. Accessed 7 May 2019.
- National Advisory Committee on Immunization (NACI). A Review of the Literature of High Dose Seasonal Influenza Vaccine for Adults 65 Years and Older. <http://www.phac-aspc.gc.ca/naci-ccni/assets/pdf/influenza-vaccine-65-plus-vaccin-contre-la-grippe-65-plus-eng.pdf>. Accessed 7 May 2019.
- US Centers for Disease Control and Prevention. Flu Vaccination Coverage, United States, 2012–13 Influenza Season. <https://www.cdc.gov/flu/fluview/coverage-1213estimates.htm>. Accessed 7 May 2019.
- US Centers for Disease Control and Prevention. Flu Vaccination Coverage, United States, 2011–12 Influenza Season. <https://www.cdc.gov/flu/fluview/coverage-1112estimates.htm>. Accessed 1 Oct 2019.
- US Centers for Disease Control and Prevention. Seasonal Influenza Vaccine Effectiveness, 2012–2013. <https://www.cdc.gov/flu/vaccines-work/2012-2013.html>. Accessed 7 May 2019.
- Cauchemez S, Carrat F, Viboud C, Valleron A, Boelle P. A bayesian mcmc approach to study transmission of influenza: application to household longitudinal data. *Stat Med*. 2004;23(22):3469–87.
- van den Driessche P. Reproduction numbers of infectious disease models. *Infect Dis Model*. 2017;2(3):288–303.
- Halloran ME, Struchiner CJ, Longini Jr IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *Am J Epidemiol*. 1997;146(10):789–803.
- Arino J, Bauch C, Brauer F, Driedger SM, Greer AL, Moghadas SM, Pizzi NJ, Sander B, Tuite A, Van Den Driessche P, et al. Pandemic influenza: modelling and public health perspectives. *Math Bio Eng*. 2011;8(1):1–20.
- Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC infect dis*. 2014;14(1):480.
- Leung NH, Xu C, Ip DK, Cowling BJ. The fraction of influenza virus infections that are asymptomatic: a systematic review and meta-analysis. *Epidemiology*. 2015;26(6):862.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.