

RESEARCH

A Comparison of the Test-Negative and the Traditional Case-Control Study Designs for Estimation of Influenza Vaccine Effectiveness under Nonrandom Vaccination

Meng Shi¹, Qian An¹, Kylie E. C. Ainslie¹, Michael Haber^{1*} and Walter A Orenstein²

*Correspondence:

mhaber@emory.edu

¹Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, 30322 Atlanta, USA

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

Abstract

Background: As annual influenza vaccination is recommended for all U.S. persons aged 6 months or older, it is unethical to conduct randomized clinical trials to estimate influenza vaccine effectiveness (VE). Observational studies are being increasingly used to estimate VE. We developed a probability model for comparing the bias and the precision of VE estimates from two case-control designs: the traditional case-control (TCC) design and the test-negative (TN) design. In both study designs, acute respiratory illness (ARI) patients seeking medical care testing positive for influenza infection are considered cases. In the TN design, ARI patients seeking medical care who test negative serve as controls, while in the TCC design, controls are randomly selected individuals from the community who did not contract an ARI.

Methods: Our model assigns each study participant a covariate corresponding to the person's health status. The probabilities of vaccination and of contracting influenza and non-influenza ARI depend on health status. Hence, our model allows non-random vaccination and confounding. In addition, the probability of seeking care for ARI may depend on vaccination and health status. We consider two outcomes of interest: symptomatic influenza (SI) and medically-attended influenza (MAI).

Results: If vaccination does not affect the probability of non-influenza ARI, then VE estimates from TN studies usually have smaller bias than estimates from TCC studies. We also found that if vaccinated influenza ARI patients are less likely to seek medical care than unvaccinated patients because the vaccine reduces symptoms' severity, then estimates of VE from both types of studies may be severely biased when the outcome of interest is SI. The bias is not present when the outcome of interest is MAI.

Conclusions: The TN design produces valid estimates of VE if (a) vaccination does not affect the probabilities of non-influenza ARI and of seeking care against influenza ARI, and (b) the confounding effects resulting from non-random vaccination are similar for influenza and non-influenza ARI. Since the bias of VE estimates depends on the outcome against which the vaccine is supposed to protect, it is important to specify the outcome of interest when evaluating the bias.

Keywords: case-control study; test-negative study; probability model; symptomatic influenza; medically-attended influenza

Background

Influenza vaccine effectiveness (VE) has to be re-estimated in every season because predominant influenza virus types, subtypes and phenotypes change from one season to the next, necessitating a new vaccine targeting different strains in most seasons. As annual influenza vaccination is now widely recommended, randomized clinical trials for estimating VE are no longer ethical in many populations, and observational studies based on patients seeking medical care for acute respiratory illnesses (ARI) are the most efficient, and hence most widely used option. However, observational studies for estimating VE are prone to multiple sources of bias.

In this paper we present a new probability model for comparing the bias and precision of VE estimates from two popular case-control study designs under nonrandom vaccination, i.e., vaccination probabilities may depend on a covariate. In both study designs, ARI patients seeking medical care who test positive for influenza infection are considered cases. In the test-negative (TN) design, ARI patients seeking medical care who test negative for influenza infection serve as controls, while in the traditional case-control (TCC) design, controls are randomly selected individuals who did not contract an ARI, usually from the same community from which the cases came. The TN design was introduced in 2007 [1], and most of the influenza VE case-control studies conducted since then have used this study design. However, TCC studies are still being used occasionally [2–4]. TCC studies are usually costlier and more resource intensive due to the need to recruit controls through a separate mechanism.

Estimates of VE from case-control studies may be subject to the following sources of bias:

(a) Probabilities of non-influenza ARI may depend on vaccination status: In TN studies, individuals with non-influenza ARI serve as controls. Therefore, TN studies may produce biased estimates of VE unless vaccinees and non-vaccinees are equally likely to develop non-influenza ARI. The validity of this assumption has not yet been confirmed. De Serres *et al.* [5] used data from randomized clinical trials to argue that this assumption is usually satisfied. However, a randomized influenza vaccine trial [6] found that vaccinees had a significantly increased risk of virologically-confirmed non-influenza infection (that may lead to ARI) as compared to those who received the placebo.

(b) Probabilities of influenza and non-influenza ARIs may depend on confounders: Covariates such as health status, age, exposure, education and socio-economic status may be associated with both the likelihood of being vaccinated and the likelihood of developing influenza and non-influenza ARIs.

(c) Vaccination may affect probability of seeking medical care in influenza patients: Several studies [7–9] suggest that vaccinated individuals who contract influenza may have milder symptoms than unvaccinated influenza patients, and therefore may be less likely to seek medical care. This effect of vaccination is not expected to change health-care-seeking behavior of non-influenza ARI patients.

(d) Probabilities of seeking medical care against ARIs may depend on confounders: Since only ARI patients who seek medical care may be included in a TNC study, and only influenza patients who seek care may be included as cases in TCC studies, covariates that are associated with both the likelihood of being

vaccinated and the likelihood of seeking care against ARI may contribute to the bias of influenza VE estimates.

(e) Misclassification bias: Diagnostic tests for influenza viruses are not 100% sensitive and specific. Vaccination status may also be misclassified.

In this work we consider the first four sources of bias. To focus on these sources, we ignore misclassification biases which are known to result in negative bias (i.e., bias toward lower estimation of VE) and are common to all studies that rely on results of diagnostic tests.

The goal of this article is to evaluate and compare the bias and precision of estimates of VE resulting from TN and TCC studies. As we will see, the bias of VE estimates may depend on the outcome of interest, i.e., the outcome against which the vaccine is expected to protect. We consider two outcomes of interest, symptomatic influenza (SI) and medically-attended influenza (MAI). In both the TN and TCC study designs, only influenza patients seeking medical care are considered cases. Therefore, one expects these study designs to produce estimates of VE against MAI. However, the media usually reports VE estimates from these case-control studies as ‘vaccine effectiveness against influenza’, without including the ‘medically-attended’ clause. As a result, the public may interpret these estimates as the effectiveness of the vaccine against *any influenza illness*, i.e., VE against SI. One of the objectives of this work is to highlight the importance of (a) clearly specifying the outcome against which the vaccine is supposed to protect, and (b) understanding that the bias of a VE estimate may be different from the two outcomes of interest.

We will (a) evaluate the bias of each of the VE estimates for each of the outcomes by comparing the expected value of the estimate with the true VE, and (b) evaluate the standard errors of the VE estimates. To perform these evaluations and comparisons, we developed a detailed stepwise probability model of the process involved in collecting data in these studies and deriving VE estimates. The model includes a covariate, health status, that may be associated with both the likelihood of being vaccinated and the propensity of seeking medical care against ARI. This allows us to assess the effects of nonrandom vaccination on the bias of VE estimates.

Methods

We first describe the real-life process involved in conducting the two types of case-control studies and obtaining the estimates of VE. We then describe the model we developed to mimic this process.

The study population

The source population for both types of case-control studies consists of all individuals receiving most of their medical care at a single clinic or at a specific network of clinics. Since influenza VE varies by age, we can assume that the model pertains to a subpopulation corresponding to a single age group.

The study designs

When a member of the study population develops an ARI, s/he may decide to report to a clinic for treatment. At the clinic, the health care provider may ask the person to be tested for influenza viruses. If the person agrees then a swab is

taken and sent to a laboratory for testing. In both study designs, a person who tests positive is eligible to be considered a case. In a TN study, an individual who tests negative is eligible to be considered a control. In a TCC study, controls are randomly selected members of the study population who have not developed ARI prior to their inclusion in the study. Usually, one or more controls are selected right after a case is identified. In both study designs, the vaccination status of every case or control is determined from manual or electronic records, or from oral histories.

Outcome of interest and true VE

In this work we evaluate estimates of VE when the outcome of interest is either SI or MAI. SI is sometimes called ‘influenza illness’ or ‘influenza ARI’. Surveillance for SI is needed in the entire study population, and for persons ill with compatible illnesses, samples of influenza are taken for verification. A person is considered a true case of SI if s/he has an ARI and is infected by an influenza virus. For MAI, a true case is defined as a person who is influenza-infected, develops an ARI, and seeks medical care. In both cases, the true VE is defined as one minus the ratio of the probability of the outcome of interest in vaccinees and non-vaccinees.

Estimation of VE and bias of VE estimates

In this work we focus on identifying the main sources of bias and their effects on the performance of the VE estimates. Some of these biases can be adjusted for in the analysis, but this is beyond the scope of the current work. In case-control studies, VE is usually estimated as one minus the odds ratio (OR) of being vaccinated in cases vs. controls. The bias of the estimate is defined as the difference between the expectation of the estimated VE and the true VE.

The model

The model we developed for comparing the estimates from the two study designs follows the scheme described above with a few simplifications. We assume that (a) when a person seeks medical care for ARI then her/his probability of being tested for influenza viruses does not depend on vaccination status, health status, or on the actual cause of ARI (influenza/non-influenza); (b) given a person’s symptoms and influenza infection status, the sensitivity and specificity of the test do not depend on the tested person’s vaccination or health status; (c) a person’s vaccination status is determined without error; and (d) controls in a TCC study are selected at random from all asymptomatic individuals in the study population (See section: The study population).

Our model includes a covariate, health status, and we assume that a person’s probabilities of being vaccinated, developing an ARI, and seeking medical care against ARI may be associated with her/his health status. In this way, the model generates possible confounding effects linking vaccination status with the probabilities of being included in the study and of becoming a case or a control.

The model consists of five steps, where the value of a single variable is determined at each step. The probability distribution of this variable may depend on the values of the variables from the previous steps. Below we define the five steps, the associated variables, and the probabilities determining each variable’s distribution.

Step 1: Health Status.

A person can be classified as “healthy” or “frail”. Define a binary variable X , where $X = 1$ for a “healthy” person and $X = 0$ for a “frail” person. Denote $\pi = P(X = 1)$.

Step 2: Vaccination.

A person may be vaccinated against influenza. Define a binary variable V , where $V = 1$ for a vaccinated person. The probability of being vaccinated may depend on health status; therefore, denote $\alpha_x = P(V = 1|X = x)$, $x = 0, 1$.

Step 3: Influenza infection and ARI.

During the influenza season, a person may become infected with an influenza virus and develop an ARI. This outcome is referred to as “influenza ARI” (FARI), where “F” stands for flu. A person may also develop an ARI not resulting from influenza infection. This outcome is referred to as “non-influenza ARI” (NFARI). We therefore define an outcome variable Y with 3 categories as follows: $Y = 0$ for no ARI, $Y = 1$ for NFARI, and $Y = 2$ for FARI. The distribution of Y depends on the person’s vaccination status, V , and health status, X . We denote $\beta_{vx} = P(Y = 1|V = v, X = x)$, $v = 0, 1$, $x = 0, 1$ and $\gamma_{vx} = P(Y = 2|V = v, X = x)$ for $v = 0, 1$, $x = 0, 1$ with $\beta_{vx} + \gamma_{vx} \leq 1$ for all v, x . Here we assume the “leaky vaccine” model, in which the vaccine provides a reduction in the probability of influenza transmission to the vaccinated person, rather than complete immunity [10]. Under this model, a vaccinee has a lower probability of becoming infected than a non-vaccinee, but is not rendered completely immune from influenza infection.

Step 4: Seeking medical care for ARI.

A person with ARI may seek medical care and, in this case, be tested for influenza viruses. We define a binary variable M with $M = 1$ for a person seeking medical care for her/his ARI. The probability of this event depends on Y (only individuals with ARI seek medical care), and it may be different for FARI and NFARI patients. In addition, the conditional distribution of M given Y may depend on X and V . We therefore define $\delta_{yvx} = P(M = 1|Y = y, V = v, X = x)$, where $y = 1, 2$, $v = 0, 1$ and $x = 0, 1$.

In order to reduce the number of parameters, we make two simplifying assumptions regarding the probabilities of seeking medical care: (1) the effect of health status on probability of seeking medical care does not depend on vaccination status or type of ARI; (2) the effect of vaccination status on probability of seeking medical care does not depend on health status (but it may depend on type of ARI).

Define a “standard person” as a person with $X = 0$ and $V = 0$. For a “standard person”, we define δ_{SN} , δ_{SF} as follows:

- $\delta_{SN} = P(M = 1|Y = 1, V = 0, X = 0) = \delta_{100}$
- $\delta_{SF} = P(M = 1|Y = 2, V = 0, X = 0) = \delta_{200}$

In addition, we define two multipliers:

- λ = multiplier for $x = 1$; λ does not depend on V and Y .
- Ψ_F = multiplier for $v=1$ only when $y=2$; Ψ_F does not depend on X .

λ is the ratio of the probabilities of seeking medical care comparing a healthy and a frail person. Ψ_F is the ratio of the probabilities of seeking care comparing a vaccinated and unvaccinated influenza ARI patient.

Then, $\{\delta_{yvx}\}$ can be written in terms of δ_{SN} , δ_{SF} and the multipliers λ , Ψ_F as follows:

- $\delta_{100} = \delta_{SN}$, $\delta_{101} = \delta_{SN} * \lambda$, $\delta_{110} = \delta_{SN}$, $\delta_{111} = \delta_{SN} * \lambda$.
- $\delta_{200} = \delta_{SF}$, $\delta_{201} = \delta_{SF} * \lambda$, $\delta_{210} = \delta_{SF} * \Psi_F$, $\delta_{211} = \delta_{SF} * \lambda * \Psi_F$.

Note: The multiplier Ψ_F reflects the effect of severity of ARI in an influenza infected person. We assume that vaccination may reduce severity of symptoms, hence a vaccinated influenza patient may be less likely to seek care than an unvaccinated patient.

Step 5: Testing for influenza infection.

Although only individuals who seek medical care for an ARI are tested for influenza infection, it will be convenient to define a binary variable T as the (possibly unobserved) test result for any person with an ARI, regardless of whether or not s/he is actually tested. Define $T = 1$ ($T = 0$) if a person would test positive (negative) for influenza if tested. Because of assumption (b) above, the probability of testing positive given the person's influenza infection status does not depend on X , V , or M . Denote $\tau_y = P(T = 1|Y = y)$ for $y = 1, 2$. Note that τ_1 is one minus the test's specificity and τ_2 is the test's sensitivity. In this study, we assume the test has 100% sensitivity and 100% specificity, i.e. $P(T = 1|Y = 1) = \tau_1 = 0$ and $P(T = 1|Y = 2) = \tau_2 = 1$.

Figure 1 shows the directed acyclic graph (DAG) of the model. Recent papers by Sullivan *et al.* [11] and Lipsitch *et al.* [12] discuss the use of DAGs to explore sources of bias of VE estimates from TN studies. A summary of the variables and parameters in our model is given in Table 1.

True VE in our model

When we evaluate the true VE, we assume that vaccination is done *at random*, i.e. for true VE we assume that vaccination status does not depend on health status X ($\alpha_0 = \alpha_1 = \alpha$).

The true VE against SI is:

$$VET_{SI} = 1 - RRT_{SI} \quad \text{where} \quad RRT_{SI} = \frac{P(Y = 2|V = 1)}{P(Y = 2|V = 0)}.$$

The true VE against MAI is:

$$VET_{MAI} = 1 - RRT_{MAI} \quad \text{where} \quad RRT_{MAI} = \frac{P(Y = 2, M = 1|V = 1)}{P(Y = 2, M = 1|V = 0)}.$$

Using the parameters defined above, VET_{SI} and VET_{MAI} can be written as:

$$VET_{SI} = 1 - RRT_{SI} = 1 - \frac{\gamma_{10}(1 - \pi) + \gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \gamma_{01}\pi}$$

$$VET_{MAI} = 1 - RRT_{MAI} = 1 - \frac{\Psi_F[\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi]}{\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi}.$$

The proofs of these results can be found in Appendix 1.

Estimates of VE in our model

In both the TN and TCC study designs, VE is estimated as one minus the odds ratio (OR) in the $C \times V$ table cross-classifying the individuals included in the study, where C is a binary indicator of case/control status with $C = 1$ for a case. For convenience, the TN and TCC studies will be represented by the letters A and B, respectively. In a TN study, the case/control variable is denoted C_A , where $(C_A = 1) = (M = 1, T = 1)$ and $(C_A = 0) = (M = 1, T = 0)$. Then the estimate of VE is: $VE_A = 1 - OR_A$, where

$$OR_A = \frac{P(C_A = 1, V = 1|M = 1)P(C_A = 0, V = 0|M = 1)}{P(C_A = 1, V = 0|M = 1)P(C_A = 0, V = 1|M = 1)}.$$

Note that all the probabilities condition on $M = 1$ as only individuals who seek medical care for ARI are included in the TN study.

In a TCC study, the case/control variable is denoted C_B . Cases are defined in the same way as in the TN study, i.e., $(C_B = 1) = (M = 1, T = 1) = (C_A = 1)$. Controls are individuals included in a random sample drawn from all the asymptomatic individuals in the study population. In other words, $(C_B = 0)$ is a random subset of $(Y = 0)$. In addition, we define a binary variable B indicating whether or not a person is included in the TCC study, i.e., $(B = 1) = (C_B = 1 \text{ or } C_B = 0)$. The VE estimate is based on the OR in the $C_B \times V$ table when all the probabilities condition on $B = 1$: $VE_B = 1 - OR_B$, where

$$OR_B = \frac{P(C_B = 1, V = 1|B = 1)P(C_B = 0, V = 0|B = 1)}{P(C_B = 1, V = 0|B = 1)P(C_B = 0, V = 1|B = 1)}.$$

Note that in a real-life study, the odds ratios are estimated from the relative frequencies of the corresponding events, rather than from their (unknown) probabilities. Therefore, the model-based estimates of VE defined above are actually the expected values of the observed estimates. For convenience we will continue to refer to them as “the VE estimates”.

Using the parameters defined above, VE_A and VE_B can be written as follows:

$$VE_A = 1 - \frac{\Psi_F[\gamma_{10}\alpha_0(1-\pi) + \lambda\gamma_{11}\alpha_1\pi][\beta_{00}(1-\alpha_0)(1-\pi) + \lambda\beta_{01}(1-\alpha_1)\pi]}{[\gamma_{00}(1-\alpha_0)(1-\pi) + \lambda\gamma_{01}(1-\alpha_1)\pi][\beta_{10}\alpha_0(1-\pi) + \lambda\beta_{11}\alpha_1\pi]},$$

$$VE_B = 1 - \frac{\Psi_F[\gamma_{10}\alpha_0(1-\pi) + \lambda\gamma_{11}\alpha_1\pi][(1-\gamma_{00}-\beta_{00})(1-\alpha_0)(1-\pi) + (1-\gamma_{01}-\beta_{01})(1-\alpha_1)\pi]}{[\gamma_{00}(1-\alpha_0)(1-\pi) + \lambda\gamma_{01}(1-\alpha_1)\pi][(1-\gamma_{10}-\beta_{10})\alpha_0(1-\pi) + (1-\gamma_{11}-\beta_{11})\alpha_1\pi]}.$$

The proofs can be found in Appendix 2.

Bias and standard errors of estimates

The bias of an estimate of VE is the difference between the expected value of the estimate and the true VE. As the true VE depends on the outcome of interest (SI or MAI), the bias of each estimated VE will be evaluated separately for each of the two outcomes.

In Appendix 3 we use approximations based on the “Delta method” for the standard errors (SEs) of odds ratios [13] to derive expressions for the SEs of both VE estimates in terms of the parameters and the corresponding sample size(s). For evaluating the SEs we consider the observed odds ratios, where the probabilities are replaced by the corresponding observed relative frequencies.

The values of bias reported in the text and tables represent absolute numbers. For example, if the true VE is 60% (i.e., 0.6) and the range of bias (-0.40, -0.20). This means that the estimated VE varies from 0.20 (underestimating the true VE = 0.6 by 0.40) to 0.80 (overestimating the true VE by 0.20).

Probability Ratios

Next, we define a few probability ratios comparing vaccinees and non-vaccinees or healthy and frail individuals. These ratios will be helpful in the presentation of the results (see Table 1 for a full list of the notations used in this paper).

- $\rho_\beta = \frac{\beta_{1x}}{\beta_{0x}}$, the ratio of the probabilities of NFARI comparing a vaccinated and an unvaccinated person of the same health status.
- $\eta_\beta = \frac{\beta_{v1}}{\beta_{v0}}$, the ratio of the probabilities of NFARI comparing a healthy and a frail person of the same vaccination status.
- $\rho_\gamma = \frac{\gamma_{1x}}{\gamma_{0x}}$, the ratio of the probabilities of FARI comparing a vaccinated and an unvaccinated person of the same health status.
- $\eta_\gamma = \frac{\gamma_{v1}}{\gamma_{v0}}$, the ratio of the probabilities of FARI comparing a healthy and a frail person of the same vaccination status.

The parameters λ and Ψ_F defined earlier are also probability ratios:

- $\lambda = \frac{\delta_{yv1}}{\delta_{yv0}}$ The ratio of the probabilities of seeking medical care comparing a healthy and a frail person of the same vaccination status. We assume that this ratio is the same for FARI and NFARI patients.
- $\Psi_F = \frac{\delta_{21x}}{\delta_{20x}}$ The ratio of the probabilities of seeking medical care comparing a vaccinated and an unvaccinated FARI patient of the same health status.

Table 2 presents the main sources of bias that can be identified from our model. The absence of bias A is essential for the validity of the TN design, since the VE estimate from this design is based on comparing the odds of being vaccinated in FARI patients (cases) and NFARI patients (controls). This bias may be a result of virus interference [6] (if vaccinees are more likely than non-vaccinees to contract NFARI, then the estimated VE will be falsely high). Biases B1 and B2 represent the effects of health status on the probabilities of NFARI and FARI, respectively. These effects, which are sometimes called the ‘*healthy vaccinee effect*’, represent the confounding resulting from association of health status with the probability of exposure (vaccination) and the outcome. Bias BS is a special case of $B1 \cap B2$. It results when health status affects both the probabilities of FARI and NFARI but the risk ratios comparing a healthy and a frail person are the same for the both types of ARIs. Bias C represents the effect of vaccination status on the probability of seeking care in patients with SI. This effect may be due to less severe symptoms

in vaccinated persons compared to unvaccinated ones. As stated earlier, we assume perfect sensitivity and specificity of the influenza test ($\tau_1 = 0$, $\tau_2 = 1$), as it is well-known that misclassifications result in negatively-biased estimates of effectiveness.

Results

Sources of Bias

We first state conditions for the unbiasedness of the VE estimate based on the TN design. The proofs of these results can be found in Appendix 4.

- (1) Under random vaccination ($\alpha_0 = \alpha_1$), the estimate of VE when the outcome of interest is SI is unbiased if biases A and C are absent. When the outcome of interest is MAI, the estimate of VE is unbiased if bias A is absent.
- (2) Under non-random vaccination ($\alpha_0 \neq \alpha_1$), the estimate of VE when the outcome of interest is SI is unbiased if biases A, B1, B2, and C are absent. When the outcome of interest is MAI, the estimate of VE is unbiased if biases A, B1, and B2 are absent.

It is interesting to note that the absence of any source of bias, the OR-based VE estimate from a TN study is unbiased even if the ‘rare disease’ assumption is not satisfied, while the OR-based estimate from a TCC study is biased. To show this, let’s use the following simplified notation: α = probability of being vaccinated, β = probability of NFARI, γ_0 and γ_1 = probabilities of FARI in unvaccinated and vaccinated, respectively, and δ = probability of seeking care. Then the true VE is $1 - \rho$, where $\rho = \gamma_1 / \gamma_0$ is the risk ratio. In a TN study, the probabilities of vaccinated and unvaccinated cases are $\alpha * \delta * \gamma_1$, and $(1 - \alpha) * \delta * \gamma_0$, respectively. The corresponding probabilities of controls are $\alpha * \delta * \beta$, and $(1 - \alpha) * \delta * \beta$, respectively. Then the OR in the table of case-control status by vaccination status equals to ρ , i.e. the true risk ratio, implying that the estimated VE is unbiased. In a TCC studies, the probabilities of cases are the same as in the TN study, while the probabilities of vaccinated and unvaccinated controls (individuals without ARI) are $\phi * \alpha * (1 - \gamma_1 - \beta)$ and $\phi * (1 - \alpha) * (1 - \gamma_0 - \beta)$, respectively, where ϕ is the sampling fraction of controls. Hence, the OR in the TCC study is $[\rho * (1 - \gamma_0 - \beta)] / (1 - \rho * \gamma_0 - \beta)$. This OR is less than ρ (the true RR) if $\rho > 0$, hence the estimated VE exceeds the true VE in a TCC study as long as the true VE is positive.

Next we explore the magnitude of the effects of various sources of bias and their combinations. We consider three scenarios for vaccination probabilities (see Table 3). In Table 4 we present the range and the maximum absolute value of the bias of VE estimates resulting from TN and TCC studies under the three vaccination scenarios and various combinations of sources of bias. For these results we used the following baseline values of some of the parameters: $\pi = 0.7$, $\beta_{00} = 0.2$, $\gamma_{00} = 0.1$, $\delta_{SN} = 0.2$, $\delta_{SF} = 0.3$, $\rho_\gamma = 0.4$. π is the probability of being ‘healthy’; β_{00} and γ_{00} are the probabilities of NFARI and FARI, respectively, for an unvaccinated ‘frail’ person; δ_{SN} and δ_{FN} are the probabilities of seeking medical care for NFARI and FARI, respectively, for an unvaccinated ‘frail’ person; ρ_γ is the risk ratio comparing the probability of FARI for a vaccinated and an unvaccinated person - thus, the true

VE against SI is $1 - 0.4 = 0.6$ (60%). The values of β 's, γ 's are based on data from various randomized placebo-controlled trials (see [14], Table A1), and the values of δ are based on data from several observational studies. In all the tables, figures and examples, values of VE are presented as fractions, rather than percentages.

In the calculations for tables 4 and 5, when a source of bias was present we used a reasonable range for the corresponding probability ratio. When bias A was present, ρ_β was allowed to vary from 0.5 to 2.0. For biases B1, B2, and BS, we allowed η_β and η_γ to vary between 0.5 and 1.0, since one would not expect frail persons to have lower probabilities of ARI, compared to healthy persons. For bias C, the ratio Ψ_F could vary between 0.5 to 1.0, since one would expect vaccination to reduce the probability that a person with SI will seek medical care compared to a person with ARI resulting from a different pathogen. For bias D, we let λ vary between 0.5 to 2.0.

For each combination of two or more sources of bias, we calculated the minimum, mean, and maximum of the bias and the absolute values of the bias by allowing the probability ratios that are not fixed to vary independently in the ranges specified above. For example, when biases A, B1, and B2 are absent, we used $\rho_\beta = \eta_\beta = \eta_\gamma = 1$, $0.5 \leq \Psi_F \leq 1$, $0.5 \leq \lambda \leq 2$.

Summary of Results

The impact of sources of bias

Our model allows us to evaluate the impact of the sources of bias listed in Table 2. Each source of bias is a result of a possible effect of vaccination or health status on the probability of FARI or NFARI or seeking care. Below we summarize our results for each of the sources of bias. We also use numerical examples to illustrate the magnitude and direction associated with each source of bias. *Unless otherwise specified, the true VEs against SI and MAI are 0.6 (60%).* In each of these examples we assume that only one source of bias is present.

- (1) Vaccination affects the probability of NFARI (bias A)
 - This bias does not depend on vaccination scenario nor on the outcome of interest (SI or MAI).
 - Estimates of VE from TN studies may suffer from severe bias.
 - This effect also affects the bias of VE estimates from TCC studies, though to a lesser extent.
 - **Example:** As the ratio of the probability of NFARI comparing vaccinated and unvaccinated persons varies from 0.5 to 2.0, VE estimates from TN studies range from 0.2 to 0.8, respectively, while VE estimates from TCC studies range from 0.67 to 0.50, respectively (Figure 2).
- (2) Health status affects the probabilities of FARI and NFARI (biases B1, B2 - the 'healthy vaccinee effect')
 - The bias does not depend on the outcome of interest (SI or MAI).
 - Under non-random vaccination, these effects may result in substantial bias of VE estimates from TN or TCC studies. However, this bias is usually less severe compared to the biases resulting from sources A, C and D.

- If the effect of health status on the probability of ARI is the same for FARI and NFARI, i.e., bias BS is present, then the TN-based estimates of VE are unbiased.
 - **Example:** Suppose that the probabilities of vaccination are 0.8 and 0.4 for healthy and frail persons, respectively. Consider three cases regarding the *risk ratios* $P(\text{ARI in a healthy person}) / P(\text{ARI in a frail person})$:
 - (a) When these risk ratios are 0.5 for NFARI and 0.8 for FARI, then the estimated VEs from TN and TCC studies are 0.51 and 0.67, respectively.
 - (b) When the risk ratios are 0.8 for NFARI and 0.5 for FARI, then estimated VEs from TN and TCC studies are 0.67 and 0.72, respectively.
 - (c) When the risk ratios for NFARI and FARI are equal and their common value ranges from 0.5 to 1.0, then estimated VEs from TN studies are always unbiased (i.e., they equal 0.6), while estimates from TCC studies range from 0.63 to 0.73. In Figures 3a and 3b, we set the risk ratio for NFARI to 0.75 and let the risk ratio for FARI vary between 0.5 to 1.0.
- (3) Vaccination affects the probability of seeking medical care for FARI, but it does not affect the probability of seeking care for NFARI (bias C)
- When this effect is present then the true VEs against SI and MAI may be different as the vaccine directly affects (reduces) the probability of seeking care in influenza cases, but not in controls. Thus the estimates' bias may depend on the outcome of interest.
 - If all other sources of bias are absent, the bias of VE estimates does not depend on the vaccination scenario.
 - Estimates of VE from TN or TCC studies may be severely biased when the outcome of interest is SI.
 - When the outcome of interest is MAI, estimates of VE from TN studies are unbiased, while the bias of estimates from TCC studies is usually small and is not affected by the magnitude of the effect underlying this source of bias.
 - **Example:** Let the ratio $R = P(\text{seeking medical care against FARI if vaccinated}) / P(\text{seeking medical care against FARI if unvaccinated})$ vary from 0.5 to 1.0. Then the true VE against SI remains fixed at 0.6, while the true VE against MAI varies with R from 0.8 to 0.6. The estimated VEs from TN studies equal the true VE against MAI for all values of R, while the estimated VEs from TCC studies vary from 0.82 to 0.63 (see Figure 4). For example, when $R=0.5$ then the true VE against MAI is 0.80, and the VE estimates from TN and TCC studies are 0.80 and 0.82, respectively. This translates into severe bias when the outcome of interest is SI but small bias when the outcome of interest is MAI.
- (4) Health status affects the probabilities of seeking care against FARI and NFARI (bias D)
- The bias of VE estimates does not depend on the outcome of interest (SI or MAI).
 - In the absence of other sources of bias, VE estimates from TN studies are unbiased regardless of the vaccination scenario.

- Under non-random vaccination, this effect may result in substantial bias in VE estimates from TCC studies.
- **Example:** We assume that the probabilities of seeking care do not depend on vaccination status. As the ratio of the probabilities of seeking care comparing healthy and frail individuals varies from 0.5 to 2.0, VE estimates from TN studies remain fixed at 0.6 (i.e., they are unbiased) under both random and non-random vaccination. When the probabilities of vaccination are 0.8 and 0.4 for healthy and frail persons, respectively, the VE estimates from TCC studies vary from 0.72 to 0.53 (Figures 5a and 5b).

In addition, we found that in some cases the true VEs against SI and MAI are different. Hence, the bias of VE estimates may depend on the outcome against which the vaccine is supposed to protect. For example, if the only sources of bias are BS, C, and D then, the VE estimate from TN studies is unbiased when considering effectiveness against MAI. The same estimate may overestimate the true VE against SI by 0.20 (i.e. 20%).

Comparison of the bias of VE estimates from TN and TCC studies:

- If one is concerned that vaccination may affect the probability of non-influenza ARI, then one should prefer the TCC study design. However, TCC-based VE estimates may still be biased in this case. For example, when the ratio of the probability of NFARI comparing a vaccinated and an unvaccinated person is 0.5, then the bias of VE estimate from TN study is -0.4 while the bias of VE estimate from TCC study is 0.07.
- Under non-random vaccination, effects of health status on probabilities of influenza and non-influenza ARI (the ‘*healthy vaccinee effect*’) may bias VE estimates from both study designs. In general, TN-based estimates perform slightly better than TCC-based estimates when this effect is believed to be the main source of bias. If the effect of health status is similar for FARI and NFARI then the TN design produces less biased estimates compared to the TCC design. For example, suppose the probabilities of vaccination are 0.4 and 0.8 for healthy and frail persons, respectively. When the risk ratios for NFARI and FARI are both 0.75, then the VE estimate from TN study is unbiased, while the bias of VE estimate from TCC study is 0.07.
- If one assumes that vaccination does not affect the probability of non-influenza ARI but one is concerned that vaccinated influenza patients are less likely to seek care than unvaccinated patients (because of reduced symptoms severity) then VE estimates may suffer from severe bias in both study designs when the outcome of interest is SI. In this case, the bias of TN-based estimates may be somewhat smaller than that of TCC-based estimates. This source of bias does not affect VE estimates when the outcome of interest is MAI! For example, suppose that the ratio comparing vaccinated and unvaccinated FARI cases w.r.t. the probability of seeking medical care is 0.5. When the outcome of interest is SI, then the bias of a VE estimate from TN study is 0.2 and the bias of a VE estimate from TCC study is 0.22. When the outcome of interest is MAI, then the VE estimate from a TN study is unbiased, while the bias of a VE estimate from a TCC study is 0.02.

- Under non-random vaccination, the TN study design is preferable to the TCC design if one is concerned about bias resulting from possible effect of a person's health status on her/his probability of seeking care against ARI. For example, suppose that the probabilities of vaccination are 0.8 and 0.4 for healthy and frail persons, respectively. When the ratio of the probabilities of seeking medical care comparing healthy and frail persons is 0.5, then the VE estimate from TN study is unbiased while the bias of VE estimate is 0.12.

Precision of VE estimates

Table 5 presents the standard errors of VE estimates from TN and TCC studies. From this table we conclude that:

- Non-random vaccination may reduce precision of VE estimates.
- If the probability of NFARI is associated with vaccination status, then VE estimates from TN studies are somewhat less precise compared to VE estimates from TCC studies, although, the differences in precision were small.
- If the probability of NFARI is not associated with vaccination status, then the precision of VE estimates from TN and TCC studies are similar.

Discussion

We developed a new model for the evaluation of the bias and precision of influenza estimates from case-control studies. The new model is more comprehensive than previously suggested models [5, 14–18] for the following reasons:

- It allows assessment of the impact of non-random vaccination.
- It incorporates a confounder (health status) which links vaccination status with the probabilities of ARI and of seeking medical care for these ARIs.
- By including parameters corresponding to the probabilities of seeking medical care, the model allows us to examine the effect of association of these probabilities with vaccination and health status on the bias of VE estimates.
- The model allows evaluating and comparing the precision of VE estimates.

Some of the sources of bias discussed here have been identified and addressed in earlier publications, but, to our best knowledge, none of the previous papers present a comprehensive discussion of all the possible sources of bias that may arise under a given model. In addition, the current model attributes the associations between the various factors involved in estimation of VE (vaccination, contracting influenza and non-influenza ARIs and seeking medical care) to an underlying covariate. Previously published models, including an earlier version of our model [14], included parameters representing these associations but these associations were not based on a common underlying factor. Therefore, we believe that the current results and conclusions may differ from those derived from less structured models.

Our calculations confirm earlier findings [15] that when the probability of non-influenza ARI depends on vaccination status, VE estimates from test-negative studies may be severely biased. However, even when this probability is not affected by vaccination, VE estimates from the two types of case-control studies considered in this work may suffer from substantial bias. In addition to the well-known '*healthy vaccinee effect*' (probabilities of vaccination and of ARI depend on health status), bias of VE estimates may result from heterogeneities in health-care-seeking

behaviors. Specifically, if vaccination reduces the probability that an influenza patient seeks medical care (because her/his symptoms are less severe than those of an unvaccinated influenza patient) then VE estimates from TN or TCC studies may grossly overestimate the true VE against SI. On the other hand, when the outcome of interest is MAI then the biases resulting from vaccine-related reduction in symptoms' severity are very small. Recent papers [7–9] found evidence of vaccine-associated reduction in influenza patient's symptoms severity. The effects of health-care-seeking behaviors on VE estimates from studies in which only ARI patients who seek medical care may become cases need to be further investigated.

The results of this work lead to the following conclusions:

- In general, estimates of influenza VE from case-control studies where only ARI patients seeking medical care are tested for influenza infection may suffer from severe bias, i.e. an absolute bias of 20% or more, especially when the outcome of interest is SI.
- The bias of VE estimates may depend on the outcome against which the vaccine is supposed to protect. When the outcome of interest is MAI, seeking medical care is a component of the outcome. In other words, the true VE against MAI reflects the vaccine effect on seeking medical care and on contracting influenza. This explains why true VE against MAI may differ from true VE against SI. When bias C is present, the vaccine directly affects (reduces) the probability of seeking care in influenza cases, but not in controls. As a result, VE against MAI is lower than VE against SI.
- Influenza VE estimates from TN studies are usually presented as 'VE against medically-attended influenza'. However, the media and lay persons may interpret these VE estimates simply as the protective effectiveness of vaccination against contracting influenza illness, i.e. VE against SI. Health authorities and the public should be made aware of this distinction.
- When the outcome of interest is SI, the TN design provides valid estimates (i.e., no or small bias) if the following assumptions are satisfied: (a) vaccination does not affect the probability of non-influenza ARI, (b) effects of confounding variables on the probabilities of influenza and non-influenza ARI are similar, and (c) vaccination does not affect the probabilities of seeking medical care for influenza ARI due to reduced severity of symptoms. When the outcome of interest is MAI, then only assumptions (a) and (b) are necessary for obtaining a valid VE estimate from a TN study.
- Estimates of VE from TCC studies have small bias when the outcome of interest is SI if assumptions (a) and (c) are satisfied, assumption (b) is replaced by the stronger assumption (b*) of no presence of confounding, and the additional assumption (d) that the probabilities of seeking medical care for ARI are not affected by potential confounders is satisfied. When the outcome of interest is MAI, then TCC-based estimates of VE have small bias under assumptions (a), (b*), and (d).
- It is important to collect more data on health-care-seeking behaviors of ARI patients and to study the effects of vaccination and potential confounders on these behaviors.

In summary, the test negative design produces less biased VE estimates, compared to the traditional case-control design provided that vaccination does not modify the

probability of non-influenza ARI. However, this very popular study design may still produce biased estimates of influenza VE, especially when the outcome of interest is symptomatic influenza. One can expect *monitored cohort studies*, where every study participant reporting an ARI is tested for influenza infection, to provide less biased estimates of VE against SI. In a future publication we plan to compare the bias of these cohort studies, which are much more expensive, with that of TN studies.

Our study has a few limitations:

- In order to focus on bias associated with the study designs, we ignored bias resulting from misclassification of infection and vaccination status.
- Our model does not account for the dynamics of outbreaks of influenza and other ARI-causing infections.
- We only consider unadjusted VE estimates as we tried to focus on sources of bias rather than on how one can reduce bias using standard or novel statistical techniques [19].

In the future we plan to improve the model by incorporating dynamics of the related processes. We also plan to use stochastic simulations to assess bias and precision of influenza VE estimates for other study designs (e.g. cohort studies) and to propose new study designs resulting in less biased VE estimates.

List of abbreviations

ARI: acute respiratory illness; FARI : influenza ARI; MAI: medically-attended influenza; NFARI: non-influenza ARI; TCC: traditional case-control; OR: odds ratio; SE: standard error; SI: symptomatic influenza; TN: test-negative; VE: vaccine effectiveness.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Dr. Haber and Ms. Shi wrote the manuscript. Dr. Orenstein, Dr. An and Ms. Ainslie made important revisions of portions of the text.

Funding

This research was supported by the National Institute of Allergies and Infectious Diseases of the National Institutes of Health (NIH) under Award R01AI110474, and by IPA 1110376-05 with the Centers for Disease Controls and Prevention (CDC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the CDC.

Acknowledgement

The authors wish to thank Dr. Ivo Foppa and two reviewers for helpful comment.

Ethics approval and consent to participate

Not applicable

Consent for publication

All the authors agreed to publish the manuscript.

Availability of data and material

Not applicable

Author details

¹Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, 30322 Atlanta, USA. ²Division of Infectious Diseases, Department of Medicine, School of Medicine, Emory University, 30322 Atlanta, USA.

References

- Skowronski, D.M., Masaro, C., Kwindt, T.L., Mak, A., Petric, M., Li, Y., Sebastian, R., Chong, M., Tam, T., De Serres, G.: Estimating vaccine effectiveness against laboratory-confirmed influenza using a sentinel physician network: results from the 2005-2006 season of dual a and b vaccine mismatch in Canada. *Vaccine* **25**(15), 2842–51 (2007). doi:10.1016/j.vaccine.2006.10.002
- Belongia, E.A., Kieke, B.A., Donahue, J.G., Greenlee, R.T., Balish, A., Foust, A., Lindstrom, S., Shay, D.K., Marshfield Influenza Study Group: Effectiveness of inactivated influenza vaccines varied substantially with antigenic match from the 2004-2005 season to the 2006-2007 season. *J Infect Dis* **199**(2), 159–67 (2009). doi:10.1086/595861
- Choi, W.S., Noh, J.Y., Seo, Y.B., Baek, J.H., Lee, J., Song, J.Y., Park, D.W., Lee, J.S., Cheong, H.J., Kim, W.J.: Case-control study of the effectiveness of the 2010-2011 seasonal influenza vaccine for prevention of laboratory-confirmed influenza virus infection in the Korean adult population. *Clin Vaccine Immunol* **20**(6), 877–81 (2013). doi:10.1128/CVI.00009-13
- Havers, F., Sokolow, L., Shay, D.K., Farley, M.M., Monroe, M., Meek, J., Daily Kirley, P., Bennett, N.M., Morin, C., Aragon, D., Thomas, A., Schaffner, W., Zansky, S.M., Baumbach, J., Ferdinands, J., Fry, A.M.: Case-control study of vaccine effectiveness in preventing laboratory-confirmed influenza hospitalizations in older adults, United States, 2010-2011. *Clin Infect Dis* **63**(10), 1304–1311 (2016). doi:10.1093/cid/ciw512
- De Serres, G., Skowronski, D.M., Wu, X.W., Ambrose, C.S.: The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomised placebo-controlled clinical trials. *Euro Surveill* **18**(37) (2013)
- Cowling, B.J., Fang, V.J., Nishiura, H., Chan, K.-H., Ng, S., Ip, D.K.M., Chiu, S.S., Leung, G.M., Peiris, J.S.M.: Increased risk of noninfluenza respiratory virus infections associated with receipt of inactivated influenza vaccine. *Clin Infect Dis* **54**(12), 1778–83 (2012). doi:10.1093/cid/cis307
- Castilla, J., Godoy, P., Dominguez, A., Martinez-Baz, I., Astray, J., Martin, V., Delgado-Rodriguez, M., Baricot, M., Soldevila, N., Maria Mayoral, J., Maria Quintana, J., Carlos Galan, J., Castro, A., Gonzalez-Candelas, F., Garin, O., Saez, M., Tamames, S., Pumarola, T., Influenza, C.C.C.: Influenza vaccine effectiveness in preventing outpatient, inpatient, and severe cases of laboratory-confirmed influenza. *Clin. Infect. Dis.* **57**(2), 167–175 (2013). doi:10.1093/cid/cit194
- VanWormer, J.J., Sundaram, M.E., Meece, J.K., Belongia, E.A.: A cross-sectional analysis of symptom severity in adults with influenza and other acute respiratory illness in the outpatient setting. *BMC Infect Dis* **14**, 231 (2014). doi:10.1186/1471-2334-14-231
- Deiss, R.G., Arnold, J.C., Chen, W.-J., Echols, S., Fairchok, M.P., Schofield, C., Danaher, P.J., McDonough, E., Ridoré, M., Mor, D., Burgess, T.H., Millar, E.V.: Vaccine-associated reduction in symptom severity among patients with influenza A/h3n2 disease. *Vaccine* **33**(51), 7160–7 (2015). doi:10.1016/j.vaccine.2015.11.004
- Haber, M., Longini, I.M. Jr, Halloran, M.E.: Measures of the effects of vaccination in a randomly mixing population. *Int J Epidemiol* **20**(1), 300–10 (1991)
- Sullivan, S.G., Tchetgen Tchetgen, E.J., Cowling, B.J.: Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *Am J Epidemiol* **184**(5), 345–53 (2016). doi:10.1093/aje/kww064
- Lipsitch, M., Jha, A., Simonsen, L.: Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. *Int J Epidemiol* (2016). doi:10.1093/ije/dyw124
- Agresti, A.: *Categorical Data Analysis*, 3rd edn. Wiley series in probability and statistics, vol. 792. Wiley, Hoboken, NJ (2013)
- Haber, M., An, Q., Foppa, I.M., Shay, D.K., Ferdinands, J.M., Orenstein, W.A.: A probability model for evaluating the bias and precision of influenza vaccine effectiveness estimates from case-control studies. *Epidemiol Infect* **143**(7), 1417–26 (2015). doi:10.1017/S0950268814002179
- Orenstein, E.W., De Serres, G., Haber, M.J., Shay, D.K., Bridges, C.B., Gargiullo, P., Orenstein, W.A.: Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. *Int J Epidemiol* **36**(3), 623–31 (2007). doi:10.1093/ije/dym021
- Ferdinands, J.M., Shay, D.K.: Magnitude of potential biases in a simulated case-control study of the effectiveness of influenza vaccination. *Clin Infect Dis* **54**(1), 25–32 (2012). doi:10.1093/cid/cir750
- Foppa, I.M., Haber, M., Ferdinands, J.M., Shay, D.K.: The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine* **31**(30), 3104–9 (2013). doi:10.1016/j.vaccine.2013.04.026
- Jackson, M.L., Nelson, J.C.: The test-negative design for estimating influenza vaccine effectiveness. *Vaccine* **31**(17), 2165–8 (2013). doi:10.1016/j.vaccine.2013.02.053
- Talbot, H.K., Nian, H., Chen, Q., Zhu, Y., Edwards, K.M., Griffin, M.R.: Evaluating the case-positive, control test-negative study design for influenza vaccine effectiveness for the frailty bias. *Vaccine* **34**(15), 1806–9 (2016). doi:10.1016/j.vaccine.2016.02.037

Figures

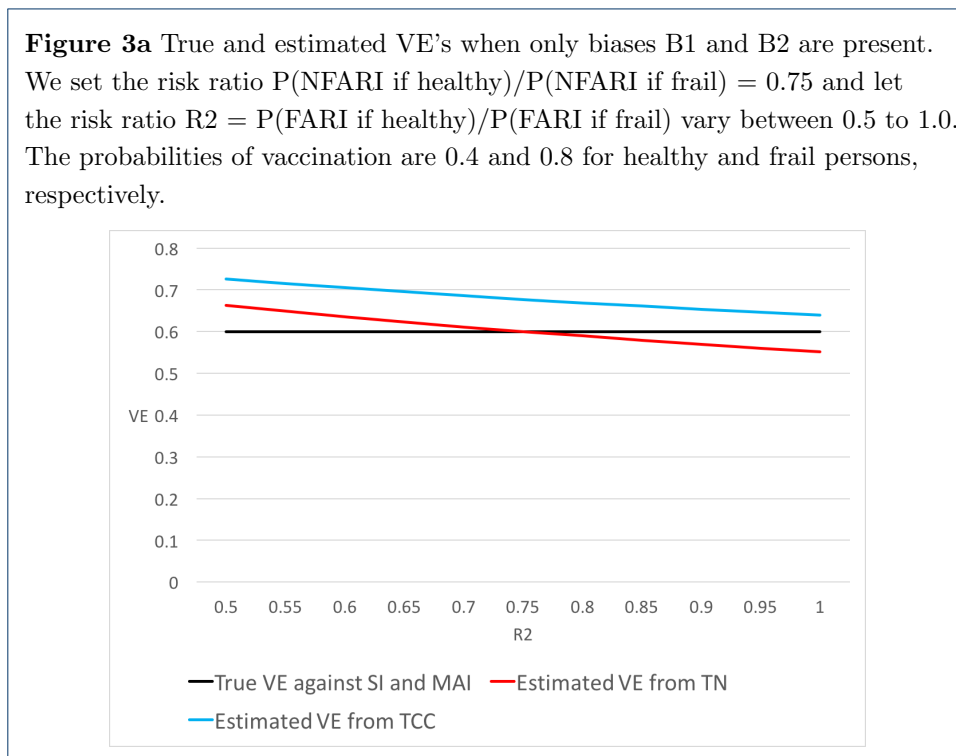
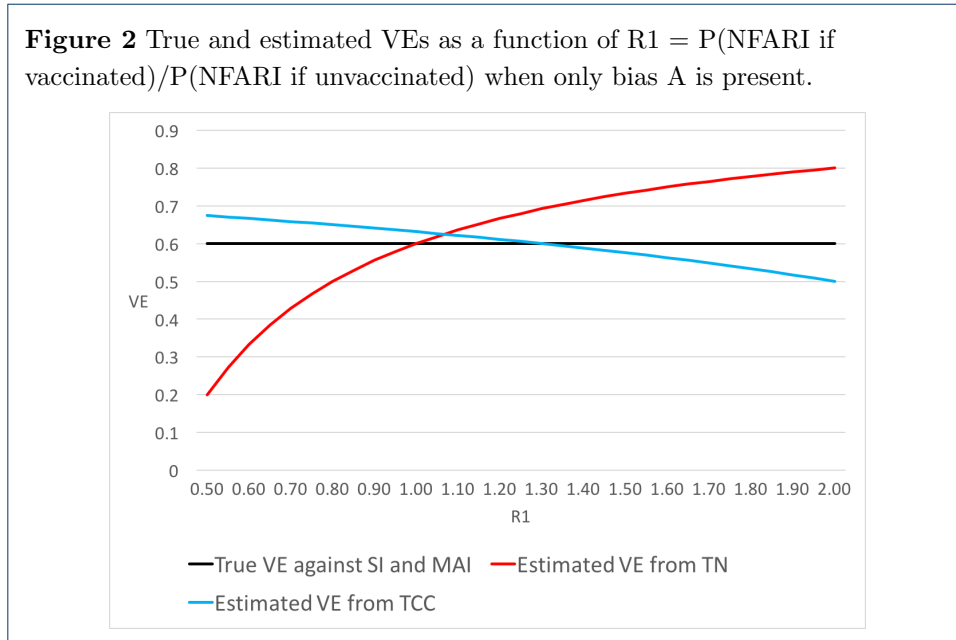
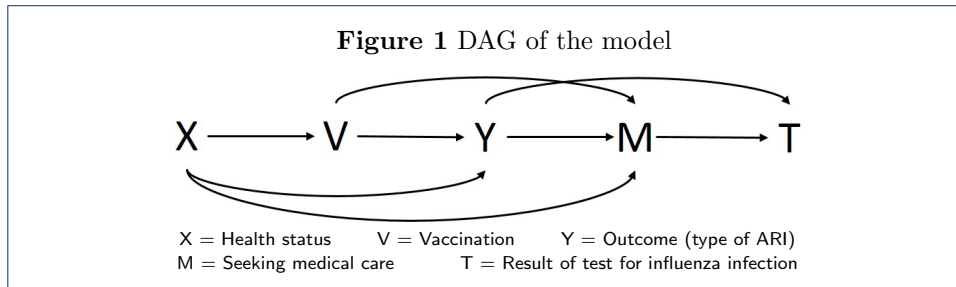


Figure 3b True and estimated VE's when only biases B1 and B2 are present. We set the risk ratio $P(\text{NFARI if healthy})/P(\text{NFARI if frail}) = 0.75$ and let the risk ratio $R2 = P(\text{FARI if healthy})/P(\text{FARI if frail})$ vary between 0.5 to 1.0. The probabilities of vaccination are 0.8 and 0.4 for healthy and frail persons, respectively.

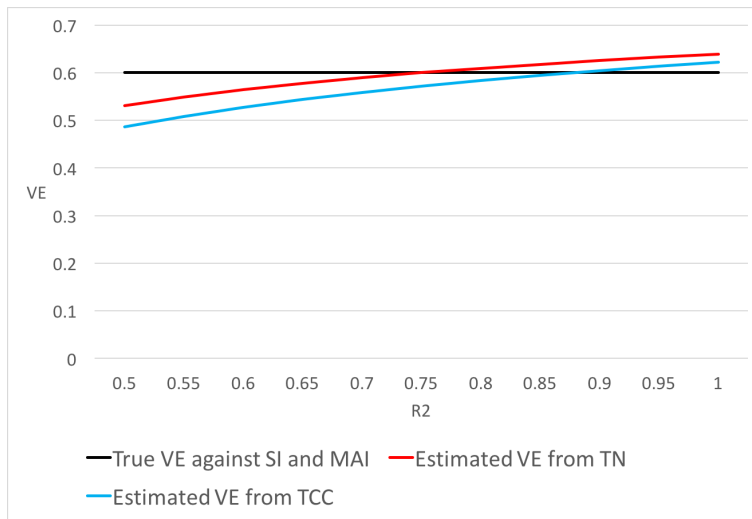


Figure 4 True and estimated VEs when only bias C is present as function of $R3 = P(\text{seeking medical care against FARI if vaccinated})/P(\text{seeking medical care against FARI if unvaccinated})$

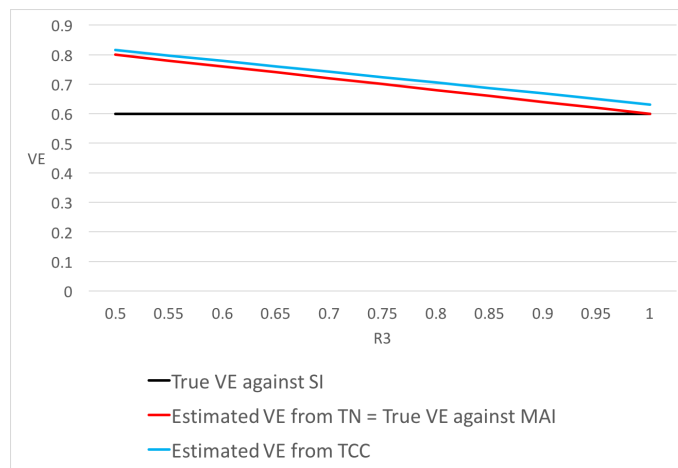


Figure 5a True and estimated VEs when only bias D is present as function of $R4 = P(\text{seeking medical care if healthy})/P(\text{seeking medical care if frail})$. Probabilities of vaccination are 0.8 and 0.4 for healthy and frail persons, respectively.

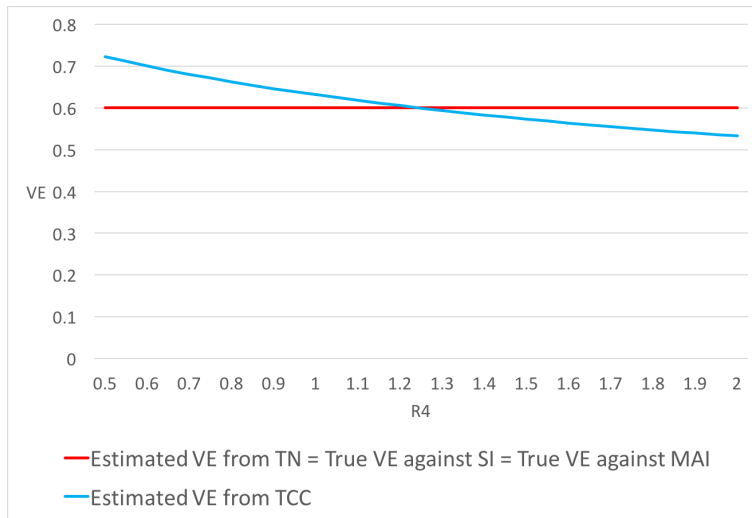
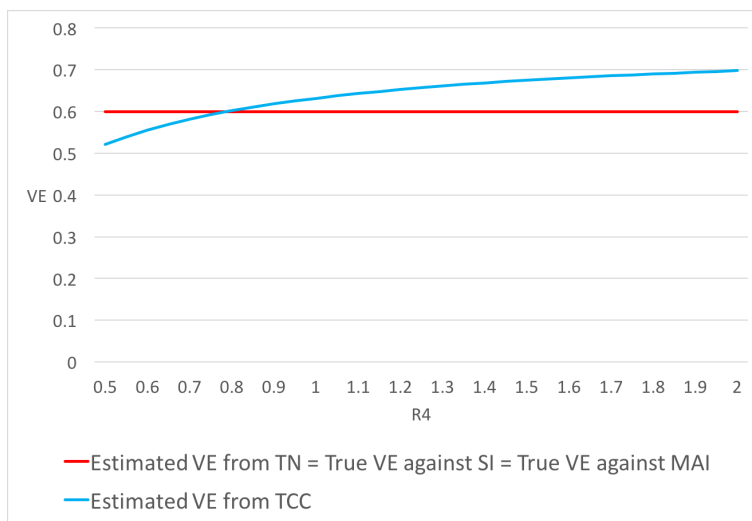


Figure 5b True and estimated VEs when only bias D is present as function of $R4 = P(\text{seeking medical care if healthy})/P(\text{seeking medical care if frail})$. Probabilities of vaccination are 0.4 and 0.8 for healthy and frail persons, respectively.



Tables

Table 1: List of parameters and other notation

Symbol	Definition	Values
X	Health status	0 - frail 1 - healthy
V	Vaccination status	0 - unvaccinated 1 - vaccinated
Y	ARI and influenza infection status	0 - no ARI 1 - NFARI 2 - FARI
M	Seeking medical care for ARI	0 - no 1 - yes
T	Result of test for influenza infection	0 - negative 1 - positive
C_A	Case/control status in TND study	0 - control 1 - case
C_B	Case/control status in TCC study	0 - control 1 - case
B	Participating in TCC study	0 - no 1 - yes
π	Probability of having better health status (i.e. healthy persons)	0.7
α_x	Probability of being vaccinated for a person of health status x	
β_{vx}	Probability of NFARI for a person of vaccination status v and health status x	
$\rho_\beta = \frac{\beta_{1x}}{\beta_{0x}}$	Ratio comparing vaccinees and non-vaccinees w.r.t. probability of NFARI	0.5-2.0
$\eta_\beta = \frac{\beta_{v1}}{\beta_{v0}}$	Ratio comparing healthy and frail persons w.r.t. probability of NFARI	0.5-1.0
γ_{vx}	Probability of FARI for a person of vaccination status v and health status x	
$\rho_\gamma = \frac{\gamma_{1x}}{\gamma_{0x}}$	Ratio comparing vaccinees and non-vaccinees w.r.t. probability of FARI	0.4 ^[1]
$\eta_\gamma = \frac{\gamma_{v1}}{\gamma_{v0}}$	Ratio comparing healthy and frail persons w.r.t. probability of FARI	0.5-1.0
δ_{SN}	Probability of seeking medical care for ARI for an unvaccinated frail person with NFARI	
δ_{SF}	Probability of seeking medical care for ARI for an unvaccinated frail person with FARI	
λ	multiplier for the probability of seeking medical care for FARI or NFARI for a healthy person	0.5-2.0
Ψ_F	multiplier for the probability of seeking medical care for FARI for a vaccinated person	0.5-1.0
τ_e	Probability that a person of illness/infection status e tests positive for influenza infection	$\tau_1 = 0, \tau_2 = 1$

[1] Assumes a true VE of 60%. Thus, the probability of FARI in a vaccinee is 40% that of a non-vaccinee.

Table 2: Sources of Bias

Source of Bias	Description
A	Vaccination affects the probability of NFARI, i.e. $\rho_\beta \neq 1$. This may result from virus interference [6].
B1	Health status affects the probability of NFARI, i.e. $\eta_\beta \neq 1$.
B2	Health status affects the probability of FARI, i.e. $\eta_\gamma \neq 1$.
BS	Health status affects the probability of FARI and NFARI, and the risk ratios comparing a healthy and a frail person are the same for both types of ARI, i.e. $\eta_\beta = \eta_\gamma \neq 1$. This is a special case of $B1 \cap B2$.
C	Vaccination affects the probability of seeking medical care for FARI, while it does not affect the probability of seeking medical care for NFARI, i.e. $\Psi_F \neq 1$. This may result from reduced severity of symptoms in vaccinated influenza patients.
D	Health status affects the probabilities of seeking medical care against FARI and NFARI, i.e. $\lambda \neq 1$.

Table 3: Three scenarios for vaccination probabilities

Vaccination Scenario	Definition
1	Random vaccination, $\alpha_0^{[1]} = \alpha_1^{[2]} = 0.6$
2	Healthy individuals are more likely to be vaccinated than frail individuals: $\alpha_0 = 0.4$, $\alpha_1 = 0.8$.
3	Healthy individuals are less likely to be vaccinated than frail individuals: $\alpha_0 = 0.8$, $\alpha_1 = 0.4$.

[1] α_0 is the probability of vaccination for frail persons;

[2] α_1 is the probability of vaccination for health persons.

Table 4: Estimate of VE against Symptomatic Influenza and Medically-Attended Influenza: Range of Bias and Maximum Absolute Value of Bias under Various Combinations of Sources of Bias

Source of Bias ^[1]	Outcome of Interest	Scenario ^[2]	Symptomatic Influenza (SI)			Medically-attended Influenza (MAI)				
			Test-Negative		Traditional Case-Control		Test-Negative		Traditional Case-Control	
			Range of Bias ^[3]	Max Value Bias ^[4]	Abs. Bias	Range of Bias	Max Value Bias	Abs. Bias	Range of Bias	Max Value Bias
None		1	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03
		2	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03
		3	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03
A		1	(-0.40, 0.20)	0.40	(-0.10, 0.07)	0.10	(-0.40, 0.20)	0.40	(-0.10, 0.07)	0.10
		2	(-0.40, 0.20)	0.40	(-0.10, 0.07)	0.10	(-0.40, 0.20)	0.40	(-0.10, 0.07)	0.10
		3	(-0.40, 0.20)	0.40	(-0.10, 0.07)	0.10	(-0.40, 0.20)	0.40	(-0.10, 0.07)	0.10
B1		1	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03
		2	(-0.13, 0.00)	0.13	(0.03, 0.05)	0.05	(-0.13, 0.00)	0.13	(0.03, 0.05)	0.05
		3	(0.00, 0.09)	0.09	(0.01, 0.03)	0.03	(0.00, 0.09)	0.09	(0.01, 0.03)	0.03
B2		1	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
		2	(0.00, 0.10)	0.10	(0.03, 0.12)	0.12	(0.00, 0.10)	0.10	(0.03, 0.12)	0.12
		3	(-0.12, 0.00)	0.12	(-0.10, 0.03)	0.10	(-0.12, 0.00)	0.12	(-0.10, 0.03)	0.10
B1,B2		1	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
		2	(-0.13, 0.10)	0.13	(0.03, 0.13)	0.13	(-0.13, 0.10)	0.13	(0.03, 0.13)	0.13
		3	(-0.12, 0.09)	0.12	(-0.12, 0.03)	0.12	(-0.12, 0.09)	0.12	(-0.12, 0.03)	0.12
BS		1	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
		2	(0.00, 0.00)	0.00	(0.03, 0.13)	0.13	(0.00, 0.00)	0.00	(0.03, 0.13)	0.13
		3	(0.00, 0.00)	0.00	(-0.12, 0.03)	0.12	(0.00, 0.00)	0.00	(-0.12, 0.03)	0.12
C		1	(0.00, 0.20)	0.20	(0.03, 0.22)	0.22	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
		2	(0.00, 0.20)	0.20	(0.03, 0.22)	0.22	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
		3	(0.00, 0.20)	0.20	(0.03, 0.22)	0.22	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
D		1	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03	(0.00, 0.00)	0.00	(0.03, 0.03)	0.03
		2	(0.00, 0.00)	0.00	(-0.07, 0.12)	0.12	(0.00, 0.00)	0.00	(-0.07, 0.12)	0.12
		3	(0.00, 0.00)	0.00	(-0.08, 0.10)	0.10	(0.00, 0.00)	0.00	(-0.08, 0.10)	0.10
C,D		1	(0.00, 0.20)	0.20	(0.03, 0.22)	0.22	(0.00, 0.00)	0.00	(0.02, 0.03)	0.03
		2	(0.00, 0.20)	0.20	(0.03, 0.22)	0.22	(0.00, 0.00)	0.00	(-0.07, 0.12)	0.12
		3	(0.00, 0.20)	0.20	(0.03, 0.22)	0.22	(0.00, 0.00)	0.00	(-0.08, 0.10)	0.10
B1,B2,C,D		1	(0.00, 0.20)	0.20	(0.02, 0.22)	0.22	(0.00, 0.00)	0.00	(0.01, 0.03)	0.03
		2	(-0.13, 0.25)	0.25	(-0.07, 0.30)	0.30	(-0.13, 0.10)	0.13	(-0.07, 0.20)	0.20
		3	(-0.13, 0.25)	0.25	(-0.30, 0.25)	0.30	(-0.13, 0.10)	0.13	(-0.30, 0.20)	0.30
BS,C,D		1	(0.00, 0.20)	0.20	(0.02, 0.22)	0.22	(0.00, 0.00)	0.00	(0.01, 0.03)	0.03
		2	(0.00, 0.20)	0.20	(-0.07, 0.30)	0.30	(0.00, 0.00)	0.00	(-0.07, 0.20)	0.20
		3	(0.00, 0.20)	0.20	(-0.30, 0.25)	0.30	(0.00, 0.00)	0.00	(-0.30, 0.10)	0.30
A,B1,B2,C,D		1	(-0.40, 0.30)	0.40	(-0.11, 0.24)	0.24	(-0.40, 0.20)	0.40	(-0.11, 0.07)	0.11
		2	(-0.67, 0.33)	0.67	(-0.23, 0.31)	0.31	(-0.67, 0.25)	0.67	(-0.23, 0.21)	0.23
		3	(-0.67, 0.33)	0.67	(-0.50, 0.27)	0.50	(-0.67, 0.11)	0.67	(-0.50, 0.13)	0.50

[1] Sources of bias: A - vaccination affects the probability of non-influenza ARI (NFARI), B1 - health status affects the probability of NFARI, B2 - health status affects the probability of influenza ARI (FARI), BS is a special case of $B1 \cap B2$ where the probabilities of FARI and NFARI depend on health status but the effect of health status on these probabilities is the same for both types of ARI, C - vaccination affects the probability of seeking medical care for FARI while it does not affect the probabilities of seeking care for NFARI, D - health status affects the probabilities of seeking medical care against FARI and NFARI.

[2] Scenario: 1 - random vaccination, 2 - healthy person more likely than frail persons to be vaccinated, 3 - healthy person less likely than frail persons to be vaccinated.

[3] Bias = estimated VE - true VE. The range of the bias is the interval between the smallest and the largest value of the bias (accounting for the sign) using different combinations of the model parameters. The sign of bias indicates the direction of the difference between the estimated and the true VE. A negative sign corresponds to underestimation while a positive bias indicates overestimation.

[4] Maximum absolute value of bias is largest difference between the estimated and the true VE when the sign of the difference is ignored:

... Little or no bias (absolute bias less than 0.05), ... Moderate bias (absolute bias greater than or equal to 0.05 and less than 0.10), ... Substantial bias (absolute bias greater than or equal to 0.10 and less than 0.20).

... Severe bias (absolute bias 0.20 or more)

Example: Under source of bias A, when the outcome of interest is SI, the TN study (under all vaccination scenarios) has a range of bias of (-0.40, 0.20). This means that the estimated VE varies from 0.20 (underestimating the true VE = 0.6 by 0.40) to 0.80 (overestimating the true VE by 0.20). When the sign of bias is ignored then the greatest difference between the estimated and the true VE is 0.40, hence the maximum absolute value of the bias is 0.40.

Table 5: Minimum, Mean and Maximum Standard Errors of VE Estimates under Various Combinations of Source of Bias

Source of Bias ^[1]	Design	Scenario ^[2]	Test-Negative			Traditional Case-Control		
			Min	Mean	Max	Min	Mean	Max
None		1	0.06	0.06	0.06	0.05	0.05	0.05
		2	0.06	0.06	0.06	0.05	0.05	0.05
		3	0.06	0.06	0.06	0.05	0.05	0.05
A		1	0.03	0.05	0.11	0.04	0.05	0.06
		2	0.03	0.05	0.10	0.04	0.05	0.06
		3	0.03	0.05	0.11	0.04	0.05	0.07
B1		1	0.05	0.05	0.06	0.05	0.05	0.05
		2	0.06	0.06	0.07	0.05	0.05	0.05
		3	0.04	0.05	0.06	0.05	0.05	0.05
B2		1	0.06	0.06	0.06	0.05	0.05	0.05
		2	0.05	0.05	0.06	0.04	0.04	0.05
		3	0.06	0.07	0.08	0.05	0.06	0.06
B1,B2		1	0.05	0.06	0.06	0.05	0.05	0.05
		2	0.05	0.06	0.07	0.04	0.04	0.05
		3	0.04	0.06	0.08	0.05	0.06	0.07
BS		1	0.06	0.06	0.06	0.05	0.05	0.05
		2	0.06	0.06	0.06	0.04	0.05	0.04
		3	0.06	0.06	0.06	0.05	0.07	0.06
C		1	0.03	0.04	0.06	0.03	0.04	0.05
		2	0.03	0.04	0.06	0.03	0.04	0.05
		3	0.03	0.05	0.06	0.03	0.04	0.05
D		1	0.06	0.06	0.06	0.05	0.05	0.05
		2	0.06	0.06	0.06	0.04	0.05	0.06
		3	0.06	0.06	0.06	0.04	0.05	0.06
C,D		1	0.03	0.04	0.06	0.03	0.04	0.05
		2	0.03	0.04	0.06	0.02	0.04	0.06
		3	0.03	0.05	0.06	0.02	0.04	0.06
B1,B2,C,D		1	0.03	0.04	0.06	0.03	0.04	0.05
		2	0.03	0.04	0.07	0.02	0.03	0.06
		3	0.02	0.05	0.08	0.02	0.04	0.09
BS,C,D		1	0.03	0.06	0.04	0.03	0.05	0.04
		2	0.03	0.06	0.04	0.02	0.06	0.03
		3	0.03	0.06	0.04	0.02	0.09	0.04
A,B1,B2,C,D		1	0.02	0.04	0.11	0.02	0.04	0.07
		2	0.02	0.04	0.14	0.01	0.04	0.08
		3	0.01	0.04	0.15	0.02	0.05	0.12

[1] Source of Bias

A	Vaccination affects the probability of NFARI
B1	Health status affects the probability of NFARI
B2	Health status affects the probability of FARI
BS	Health status affects the probability of FARI and NFARI, and the risk ratios comparing a healthy and a frail person are the same for both types of ARI.
C	Vaccination affects the probability of seeking medical care for FARI, while it does not affect the probability of seeking medical care for NFARI.
D	Health status affects the probabilities of seeking medical care against FARI and NFARI.

[2] Vaccination Scenarios

1	Random vaccination
2	Healthy person more likely than frail persons to be vaccinated
3	Healthy person less likely than frail persons to be vaccinated

Appendices

Appendix 1: True VE's in our model

The true VE against SI is:

$$VET_{SI} = 1 - RRT_{SI} \quad \text{where} \quad RRT_{SI} = \frac{P(Y = 2|V = 1)}{P(Y = 2|V = 0)}.$$

Since

$$\begin{aligned} P(Y = y|V = v) &= \sum_x P(Y = y|V = v, X = x)P(X = x|V = v) \\ P(V = v) &= \sum_x P(V = v|X = x)P(X = x) \\ P(X = x|V = v) &= \frac{P(V = v|X = x)P(X = x)}{P(V = v)} = \frac{P(V = v|X = x)P(X = x)}{\sum_x P(V = v|X = x)P(X = x)} \end{aligned}$$

we have

$$\begin{aligned} P(X = 0|V = 0) &= \frac{P(V = 0|X = 0)P(X = 0)}{\sum_x P(V = 0|X = x)P(X = x)} = \frac{(1 - \alpha_0)(1 - \pi)}{(1 - \alpha_0)(1 - \pi) + (1 - \alpha_1)\pi} = 1 - \pi, \\ P(X = 0|V = 1) &= \frac{P(V = 1|X = 0)P(X = 0)}{\sum_x P(V = 1|X = x)P(X = x)} = \frac{\alpha_0(1 - \pi)}{\alpha_0(1 - \pi) + \alpha_1\pi} = 1 - \pi, \\ P(X = 1|V = 0) &= \frac{P(V = 0|X = 1)P(X = 1)}{\sum_x P(V = 0|X = x)P(X = x)} = \frac{(1 - \alpha_1)\pi}{(1 - \alpha_0)(1 - \pi) + (1 - \alpha_1)\pi} = \pi, \\ P(X = 1|V = 1) &= \frac{P(V = 1|X = 1)P(X = 1)}{\sum_x P(V = 1|X = x)P(X = x)} = \frac{\alpha_1\pi}{\alpha_0(1 - \pi) + \alpha_1\pi} = \pi. \end{aligned}$$

Since, for true VE, we have: $\alpha_0 = \alpha_1$. We can get,

$$\begin{aligned} P(Y = 2|V = 1) &= \sum_x P(Y = 2|V = 1, X = x)P(X = x|V = 1) \\ &= P(Y = 2|V = 1, X = 0)P(X = 0|V = 1) + P(Y = 2|V = 1, X = 1)P(X = 1|V = 1) \\ &= \gamma_{10}(1 - \pi) + \gamma_{11}\pi \end{aligned}$$

$$\begin{aligned} P(Y = 2|V = 0) &= \sum_x P(Y = 2|V = 0, X = x)P(X = x|V = 0) \\ &= P(Y = 2|V = 0, X = 0)P(X = 0|V = 0) + P(Y = 2|V = 0, X = 1)P(X = 1|V = 0) \\ &= \gamma_{00}(1 - \pi) + \gamma_{01}\pi \end{aligned}$$

So that,

$$RRT_{SI} = \frac{P(Y = 2|V = 1)}{P(Y = 2|V = 0)} = \frac{\gamma_{10}(1 - \pi) + \gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \gamma_{01}\pi}$$

Therefore,

$$VET_{SI} = 1 - RRT_{SI} = 1 - \frac{\gamma_{10}(1 - \pi) + \gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \gamma_{01}\pi} \quad \text{Q.E.D.}$$

The true VE against MAI is:

$$VET_{MAI} = 1 - RRT_{MAI} \quad \text{where} \quad RRT_{MAI} = \frac{P(Y = 2, M = 1|V = 1)}{P(Y = 2, M = 1|V = 0)}.$$

Since

$$P(Y = 2, M = 1|V = v, X = x) = P(M = 1|Y = 2, V = v, X = x) * P(Y = 2|V = v, X = x)$$

$$P(Y = 2, M = 1|V = v) = \sum_x P(Y = 2, M = 1|V = v, X = x)P(X = x|V = v)$$

we can get,

$$\begin{aligned} P(Y = 2, M = 1|V = 1) &= \sum_x P(Y = 2, M = 1|V = 1, X = x)P(X = x|V = 1) \\ &= \delta_{210}\gamma_{10}(1 - \pi) + \delta_{211}\gamma_{11}\pi \\ P(Y = 2, M = 1|V = 0) &= \sum_x P(Y = 2, M = 1|V = 0, X = x)P(X = x|V = 0) \\ &= \delta_{200}\gamma_{00}(1 - \pi) + \delta_{201}\gamma_{01}\pi \end{aligned}$$

Therefore,

$$\begin{aligned} RRT_{MAI} &= \frac{P(Y = 2, M = 1|V = 1)}{P(Y = 2, M = 1|V = 0)} \\ &= \frac{\delta_{210}\gamma_{10}(1 - \pi) + \delta_{211}\gamma_{11}\pi}{\delta_{200}\gamma_{00}(1 - \pi) + \delta_{201}\gamma_{01}\pi} = \frac{\delta_{SF}\Psi_F\gamma_{10}(1 - \pi) + \delta_{SF}\Psi_F\lambda\gamma_{11}\pi}{\delta_{SF}\gamma_{00}(1 - \pi) + \delta_{SF}\lambda\gamma_{01}\pi} \\ &= \frac{\Psi_F[\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi]}{\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi} \end{aligned}$$

Hence,

$$VET_{MAI} = 1 - RRT_{MAI} = 1 - \frac{\Psi_F[\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi]}{\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi}. \quad \text{Q.E.D.}$$

Appendix 2: Model-based estimates of VE

The model-based estimate from TN study is:

$$VE_A = 1 - OR_A, \quad \text{where} \quad OR_A = \frac{P(C_A = 1, V = 1|M = 1)P(C_A = 0, V = 0|M = 1)}{P(C_A = 1, V = 0|M = 1)P(C_A = 0, V = 1|M = 1)}.$$

OR_A can be written as:

$$OR_A = \frac{P(M = 1, T = 1, V = 1)P(M = 1, T = 0, V = 0)}{P(M = 1, T = 1, V = 0)P(M = 1, T = 0, V = 1)}$$

$$\begin{aligned} P(M = 1, T = 1, V = v) &= \sum_x P(M = 1, T = 1, V = v|X = x)P(X = x) \\ &= \sum_x P(T = 1|M = 1, V = v, X = x)P(M = 1, V = v|X = x)P(X = x) \\ &= \sum_x P(Y = 2|M = 1, V = v, X = x)P(M = 1|V = v, X = x)P(V = v|X = x)P(X = x) \\ &= \sum_x P(M = 1|Y = 2, V = v, X = x)P(Y = 2|V = v, X = x)P(V = v|X = x)P(X = x) \end{aligned}$$

$$\begin{aligned} P(M = 1, T = 0, V = v) &= \sum_x P(M = 1, T = 0, V = v|X = x)P(X = x) \\ &= \sum_x P(T = 0|M = 1, V = v, X = x)P(M = 1, V = v|X = x)P(X = x) \\ &= \sum_x P(Y = 1|M = 1, V = v, X = x)P(M = 1|V = v, X = x)P(V = v|X = x)P(X = x) \\ &= \sum_x P(M = 1|Y = 1, V = v, X = x)P(Y = 1|V = v, X = x)P(V = v|X = x)P(X = x) \end{aligned}$$

So that,

$$\begin{aligned}
 P(M = 1, T = 1, V = 0) &= \delta_{200}\gamma_{00}(1 - \alpha_0)(1 - \pi) + \delta_{201}\gamma_{01}(1 - \alpha_1)\pi \\
 &= \delta_{SF}[\gamma_{00}(1 - \alpha_0)(1 - \pi) + \lambda\gamma_{01}(1 - \alpha_1)\pi] \\
 P(M = 1, T = 1, V = 1) &= \delta_{210}\gamma_{10}\alpha_0(1 - \pi) + \delta_{211}\gamma_{11}\alpha_1\pi \\
 &= \delta_{SF}\Psi_F[\gamma_{10}\alpha_0(1 - \pi) + \lambda\gamma_{11}\alpha_1\pi] \\
 P(M = 1, T = 0, V = 0) &= \delta_{100}\beta_{00}(1 - \alpha_0)(1 - \pi) + \delta_{101}\beta_{01}(1 - \alpha_1)\pi \\
 &= \delta_{SN}[\beta_{00}(1 - \alpha_0)(1 - \pi) + \lambda\beta_{01}(1 - \alpha_1)\pi] \\
 P(M = 1, T = 0, V = 1) &= \delta_{110}\beta_{10}\alpha_0(1 - \pi) + \delta_{111}\beta_{11}\alpha_1\pi \\
 &= \delta_{SN}[\beta_{10}\alpha_0(1 - \pi) + \lambda\beta_{11}\alpha_1\pi]
 \end{aligned}$$

Therefore,

$$VE_A = 1 - \frac{\Psi_F[\gamma_{10}\alpha_0(1 - \pi) + \lambda\gamma_{11}\alpha_1\pi][\beta_{00}(1 - \alpha_0)(1 - \pi) + \lambda\beta_{01}(1 - \alpha_1)\pi]}{[\gamma_{00}(1 - \alpha_0)(1 - \pi) + \lambda\gamma_{01}(1 - \alpha_1)\pi][\beta_{10}\alpha_0(1 - \pi) + \lambda\beta_{11}\alpha_1\pi]}. \quad \text{Q.E.D.}$$

The model-based estimates from TCC study is:

$$VE_B = 1 - OR_B, \quad \text{where} \quad OR_B = \frac{P(C_B = 1, V = 1|B = 1)P(C_B = 0, V = 0|B = 1)}{P(C_B = 1, V = 0|B = 1)P(C_B = 0, V = 1|B = 1)}.$$

OR_B can be written as:

$$OR_B = \frac{P(M = 1, T = 1, V = 1)P(Y = 0, V = 0)}{P(M = 1, T = 1, V = 0)P(Y = 0, V = 1)}.$$

Since,

$$\begin{aligned}
 P(Y = 0, V = v) &= \sum_x P(Y = 0, V = v|X = x)P(X = x) \\
 &= \sum_x P(Y = 0|V = v, X = x)P(V = v|X = x)P(X = x)
 \end{aligned}$$

and $P(Y = 0|V = v, X = x) = 1 - P(Y = 1|V = v, X = x) - P(Y = 2|V = v, X = x) = 1 - \gamma_{vx} - \beta_{vx}$,
so we have:

$$\begin{aligned}
 P(Y = 0, V = 0) &= (1 - \gamma_{00} - \beta_{00})(1 - \alpha_0)(1 - \pi) + (1 - \gamma_{01} - \beta_{01})(1 - \alpha_1)\pi \\
 P(Y = 0, V = 1) &= (1 - \gamma_{10} - \beta_{10})\alpha_0(1 - \pi) + (1 - \gamma_{11} - \beta_{11})\alpha_1\pi
 \end{aligned}$$

Therefore,

$$VE_B = 1 - \frac{\Psi_F[\gamma_{10}\alpha_0(1 - \pi) + \lambda\gamma_{11}\alpha_1\pi][(1 - \gamma_{00} - \beta_{00})(1 - \alpha_0)(1 - \pi) + (1 - \gamma_{01} - \beta_{01})(1 - \alpha_1)\pi]}{[\gamma_{00}(1 - \alpha_0)(1 - \pi) + \lambda\gamma_{01}(1 - \alpha_1)\pi][(1 - \gamma_{10} - \beta_{10})\alpha_0(1 - \pi) + (1 - \gamma_{11} - \beta_{11})\alpha_1\pi]}. \quad \text{Q.E.D.}$$

Appendix 3: Standard Errors of the VE Estimates

For TN study, the approximate standard error of VE_A is:

$$\begin{aligned}
 SE(VE_A) &= SE(OR_A) \approx OR_A * SE(\log(OR_A)) \\
 &\approx \frac{p_{11}^A(1 - p_{01}^A)}{p_{01}^A(1 - p_{11}^A)} \sqrt{\frac{1}{N_A} \left[\frac{1}{P_{V1}^A P_{11}^A} + \frac{1}{(1 - P_{V1}^A) P_{01}^A} + \frac{1}{P_{V1}^A(1 - P_{11}^A)} + \frac{1}{(1 - P_{V1}^A)(1 - P_{01}^A)} \right]}
 \end{aligned}$$

where N_A is the number of persons who were tested for influenza ($M=1$), i.e., the total sample size for the TN study. The probabilities ($p_{V1}^A, p_{01}^A, p_{11}^A$) can be written in terms of the parameters defined earlier.

Base on what we got earlier, we know

$$\begin{aligned}
P(M = 1, V = 1) &= \sum_t P(M = 1, V = 1, T = t) \\
&= \delta_{110}\beta_{10}\alpha_0(1 - \pi) + \delta_{111}\beta_{11}\alpha_1\pi + \delta_{210}\gamma_{10}\alpha_0(1 - \pi) + \delta_{211}\gamma_{11}\alpha_1\pi \\
&= \alpha_0(1 - \pi)[\delta_{SN}\beta_{10} + \delta_{SF}\Psi_F\gamma_{10}] + \alpha_1\pi\lambda[\delta_{SN}\beta_{11} + \delta_{SF}\Psi_F\gamma_{11}] \\
P(M = 1, V = 0) &= \sum_t P(M = 1, V = 0, T = t) \\
&= \delta_{100}\beta_{00}(1 - \alpha_0)(1 - \pi) + \delta_{101}\beta_{01}(1 - \alpha_1)\pi + \delta_{200}\gamma_{00}(1 - \alpha_0)(1 - \pi) + \delta_{201}\gamma_{01}(1 - \alpha_1)\pi \\
&= (1 - \alpha_0)(1 - \pi)[\delta_{SN}\beta_{00} + \delta_{SF}\gamma_{00}] + (1 - \alpha_1)\pi\lambda[\delta_{SN}\beta_{01} + \delta_{SF}\gamma_{01}]
\end{aligned}$$

Thus,

$$\begin{aligned}
P(M = 1) &= \sum_v P(M = 1, V = v) \\
&= (1 - \alpha_0)(1 - \pi)[\delta_{SN}\beta_{00} + \delta_{SF}\gamma_{00}] + (1 - \alpha_1)\pi\lambda[\delta_{SN}\beta_{01} + \delta_{SF}\gamma_{01}] \\
&\quad + \alpha_0(1 - \pi)[\delta_{SN}\beta_{10} + \delta_{SF}\Psi_F\gamma_{10}] + \alpha_1\pi\lambda[\delta_{SN}\beta_{11} + \delta_{SF}\Psi_F\gamma_{11}] \\
&= (1 - \pi)[(1 - \alpha_0)(\delta_{SN}\beta_{00} + \delta_{SF}\gamma_{00}) + \alpha_0(\delta_{SN}\beta_{10} + \delta_{SF}\Psi_F\gamma_{10})] \\
&\quad + \pi\lambda[(1 - \alpha_1)(\delta_{SN}\beta_{01} + \delta_{SF}\gamma_{01}) + \alpha_1(\delta_{SN}\beta_{11} + \delta_{SF}\Psi_F\gamma_{11})]
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
P_{V1}^A &= P(V = 1|M = 1) = \frac{P(V = 1, M = 1)}{P(M = 1)} \\
&= \frac{\alpha_0(1 - \pi)[\delta_{SN}\beta_{10} + \delta_{SF}\Psi_F\gamma_{10}] + \alpha_1\pi\lambda[\delta_{SN}\beta_{11} + \delta_{SF}\Psi_F\gamma_{11}]}{(1 - \pi)[(1 - \alpha_0)(\delta_{SN}\beta_{00} + \delta_{SF}\gamma_{00}) + \alpha_0(\delta_{SN}\beta_{10} + \delta_{SF}\Psi_F\gamma_{10})] + \pi\lambda[(1 - \alpha_1)(\delta_{SN}\beta_{01} + \delta_{SF}\gamma_{01}) + \alpha_1(\delta_{SN}\beta_{11} + \delta_{SF}\Psi_F\gamma_{11})]}
\end{aligned}$$

$$\begin{aligned}
p_{01}^A &= \frac{P(M = 1, T = 1|V = 0)}{P(M = 1|V = 0)} = \frac{P(M = 1, T = 1, V = 0)}{P(M = 1, V = 0)} \\
&= \frac{\delta_{200}\gamma_{00}(1 - \alpha_0)(1 - \pi) + \delta_{201}\gamma_{01}(1 - \alpha_1)\pi}{\delta_{100}\beta_{00}(1 - \alpha_0)(1 - \pi) + \delta_{101}\beta_{01}(1 - \alpha_1)\pi + \delta_{200}\gamma_{00}(1 - \alpha_0)(1 - \pi) + \delta_{201}\gamma_{01}(1 - \alpha_1)\pi} \\
&= \frac{\delta_{SF}[\gamma_{00}(1 - \alpha_0)(1 - \pi) + \lambda\gamma_{01}(1 - \alpha_1)\pi]}{(1 - \alpha_0)(1 - \pi)[\delta_{SN}\beta_{00} + \delta_{SF}\gamma_{00}] + (1 - \alpha_1)\pi\lambda[\delta_{SN}\beta_{01} + \delta_{SF}\gamma_{01}]}
\end{aligned}$$

$$\begin{aligned}
p_{11}^A &= \frac{P(M = 1, T = 1|V = 1)}{P(M = 1|V = 1)} = \frac{P(M = 1, T = 1, V = 1)}{P(M = 1, V = 1)} \\
&= \frac{\delta_{210}\gamma_{10}\alpha_0(1 - \pi) + \delta_{211}\gamma_{11}\alpha_1\pi}{\delta_{110}\beta_{10}\alpha_0(1 - \pi) + \delta_{111}\beta_{11}\alpha_1\pi + \delta_{210}\gamma_{10}\alpha_0(1 - \pi) + \delta_{211}\gamma_{11}\alpha_1\pi} \\
&= \frac{\delta_{SF}\Psi_F[\gamma_{10}\alpha_0(1 - \pi) + \lambda\gamma_{11}\alpha_1\pi]}{\alpha_0(1 - \pi)[\delta_{SN}\beta_{10} + \delta_{SF}\Psi_F\gamma_{10}] + \alpha_1\pi\lambda[\delta_{SN}\beta_{11} + \delta_{SF}\Psi_F\gamma_{11}]}
\end{aligned}$$

In the TCC study, the approximate standard error of VE_B is:

$$\begin{aligned}
SE(VE_B) &= SE(OR_B) \approx OR_B * SE(\log(OR_B)) \\
&\approx \frac{p_{11}^B(1 - p_{10}^B)}{p_{10}^B(1 - p_{11}^B)} \sqrt{\frac{1}{N_{C1}^B p_{11}^B} + \frac{1}{N_{C1}^B(1 - p_{11}^B)} + \frac{1}{N_{C0}^B p_{10}^B} + \frac{1}{N_{C0}^B(1 - p_{10}^B)}}
\end{aligned}$$

where N_{C1}^b is the number of cases and N_{C0}^b is the number of controls. The probabilities (p_{10}^B, p_{11}^B) can be written in terms of the parameters defined earlier:

$$\begin{aligned}
 p_{10}^B &= P(V = 1|C_B = 0, B = 1) = \frac{P(Y = 0, V = 1)}{P(Y = 0)} = \frac{P(Y = 0, V = 1)}{\sum_v P(Y = 0, V = v)} \\
 &= \frac{(1 - \gamma_{10} - \beta_{10})\alpha_0(1 - \pi) + (1 - \gamma_{11} - \beta_{11})\alpha_1\pi}{(1 - \pi)[(1 - \gamma_{10} - \beta_{10})\alpha_0 + (1 - \gamma_{00} - \beta_{00})(1 - \alpha_0)] + \pi[(1 - \gamma_{11} - \beta_{11})\alpha_1 + (1 - \gamma_{01} - \beta_{01})(1 - \alpha_1)]}
 \end{aligned}$$

$$\begin{aligned}
 p_{11}^B &= P(V = 1|C_B = 1, B = 1) = \frac{P(M = 1, T = 1, V = 1)}{P(M = 1, T = 1)} = \frac{P(M = 1, T = 1, V = 1)}{\sum_v P(M = 1, T = 1, V = v)} \\
 &= \frac{\delta_{210}\gamma_{10}\alpha_0(1 - \pi) + \delta_{211}\gamma_{11}\alpha_1\pi}{\delta_{200}\gamma_{00}(1 - \alpha_0)(1 - \pi) + \delta_{201}\gamma_{01}(1 - \alpha_1)\pi + \delta_{210}\gamma_{10}\alpha_0(1 - \pi) + \delta_{211}\gamma_{11}\alpha_1\pi} \\
 &= \frac{\Psi_F[\gamma_{10}\alpha_0(1 - \pi) + \lambda\gamma_{11}\alpha_1\pi]}{[\gamma_{00}(1 - \alpha_0)(1 - \pi) + \lambda\gamma_{01}(1 - \alpha_1)\pi] + \Psi_F[\gamma_{10}\alpha_0(1 - \pi) + \lambda\gamma_{11}\alpha_1\pi]}
 \end{aligned}$$

Appendix 4: Unbiasness under Random and Non-random Vaccination

Unbiasness under Random Vaccination

If the vaccination is done at random, then $\alpha_0 = \alpha_1$. The VE estimates can be written as:

$$VE_A = 1 - \frac{\Psi_F[\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi][\beta_{00}(1 - \pi) + \lambda\beta_{01}\pi]}{[\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi][\beta_{10}(1 - \pi) + \lambda\beta_{11}\pi]}$$

$$VE_B = 1 - \frac{\Psi_F[\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi][(1 - \gamma_{00} - \beta_{00})(1 - \pi) + (1 - \gamma_{01} - \beta_{01})\pi]}{[\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi][(1 - \gamma_{10} - \beta_{10})(1 - \pi) + (1 - \gamma_{11} - \beta_{11})\pi]}$$

(1) If $\rho_\beta = \Psi_F = 1$, and one of the following conditions is satisfied, then $VE_A = VET_{SI}$.

- (a) $\lambda = 1$;
- (b) $\eta_\gamma = 1$.

Proof:

Since $\rho_{\beta_x} = \Psi_F = 1$, then $\frac{\beta_{10}}{\beta_{00}} = \frac{\beta_{11}}{\beta_{01}} = \Psi_F = 1$. We have:

$$\beta_{10} = \beta_{00} \quad \text{and} \quad \beta_{11} = \beta_{01}.$$

So,

$$\begin{aligned}
 VE_A &= 1 - \frac{\Psi_F[\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi][\beta_{00}(1 - \pi) + \lambda\beta_{01}\pi]}{[\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi][\beta_{00}(1 - \pi) + \lambda\beta_{01}\pi]} \\
 &= 1 - \frac{\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi}
 \end{aligned}$$

If (a) $\lambda = 1$ is satisfied, so

$$VE_A = 1 - \frac{\gamma_{10}(1 - \pi) + \gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \gamma_{01}\pi} = VET_{SI}.$$

If (b) $\eta_\gamma = 1$ is satisfied, then $\gamma_{01} = \gamma_{00}$ and $\gamma_{11} = \gamma_{10}$. Thus,

$$VET_{SI} = 1 - \frac{\gamma_{10}(1 - \pi) + \gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \gamma_{01}\pi} = 1 - \frac{\gamma_{11}}{\gamma_{01}}.$$

Hence,

$$\begin{aligned}
 VE_A &= 1 - \frac{\gamma_{10}(1 - \pi) + \lambda\gamma_{11}\pi}{\gamma_{00}(1 - \pi) + \lambda\gamma_{01}\pi} = 1 - \frac{\gamma_{11}(1 - \pi) + \lambda\gamma_{11}\pi}{\gamma_{01}(1 - \pi) + \lambda\gamma_{01}\pi} = 1 - \frac{\gamma_{11}[(1 - \pi) + \lambda\pi]}{\gamma_{01}[(1 - \pi) + \lambda\pi]} \\
 &= 1 - \frac{\gamma_{11}}{\gamma_{01}} = VET_{SI} \quad \text{Q.E.D.}
 \end{aligned}$$

(2) If $\rho_\beta = 1$, then $VE_A = VET_{MAI}$

Proof:

Since $\rho_\beta = 1$, then $\frac{\beta_{10}}{\beta_{00}} = \frac{\beta_{11}}{\beta_{01}} = 1$.

So,

$$\begin{aligned} VEA &= 1 - \frac{\Psi_F[\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi][\beta_{00}(1-\pi) + \lambda\beta_{01}\pi]}{[\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi][\beta_{00}(1-\pi) + \lambda\beta_{01}\pi]} \\ &= 1 - \frac{\Psi_F[\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi]}{\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi} \\ &= VET_{MAI} \quad \text{Q.E.D.} \end{aligned}$$

(3) If $\lambda = 1$ and $1 - \gamma_{1x} - \beta_{1x} = \Psi_F(1 - \gamma_{0x} - \beta_{0x})$, where $x = 0, 1$, then $VE_B = VET_{SI}$

Proof:

Since $1 - \gamma_{1x} - \beta_{1x} = \Psi_F(1 - \gamma_{0x} - \beta_{0x})$, where $x = 0, 1$, and $\lambda = 1$, then

$$\begin{aligned} VEB &= 1 - \frac{\Psi_F[\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi][(1-\gamma_{00}-\beta_{00})(1-\pi) + (1-\gamma_{01}-\beta_{01})\pi]}{[\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi][\Psi_F(1-\gamma_{00}-\beta_{00})(1-\pi) + \Psi_F(1-\gamma_{01}-\beta_{01})\pi]} \\ &= 1 - \frac{\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi}{\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi} \\ &= 1 - \frac{\gamma_{10}(1-\pi) + \gamma_{11}\pi}{\gamma_{00}(1-\pi) + \gamma_{01}\pi} \\ &= VET_{SI} \quad \text{Q.E.D.} \end{aligned}$$

(4) If $\gamma_{1x} + \beta_{1x} = \gamma_{0x} + \beta_{0x}$, where $x = 0, 1$, then $VE_B = VET_{MAI}$

Proof:

Since $\gamma_{1x} + \beta_{1x} = \gamma_{0x} + \beta_{0x}$, $x = 0, 1$, then

$$1 - \gamma_{1x} - \beta_{1x} = 1 - \gamma_{0x} - \beta_{0x}, \quad \text{where } x = 0, 1$$

So:

$$\begin{aligned} VEB &= 1 - \frac{\Psi_F[\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi][(1-\gamma_{00}-\beta_{00})(1-\pi) + (1-\gamma_{01}-\beta_{01})\pi]}{[\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi][(1-\gamma_{00}-\beta_{00})(1-\pi) + (1-\gamma_{01}-\beta_{01})\pi]} \\ &= 1 - \frac{\Psi_F[\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi]}{\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi} \\ &= VET_{MAI} \quad \text{Q.E.D.} \end{aligned}$$

Unbiasness under Non-random Vaccination

If the vaccination is not done at random, then $\alpha_0 \neq \alpha_1$.

(5) If $\rho_\beta = \eta_\beta = \eta_\gamma = \Psi_F = 1$, then $VE_A = VET_{SI}$.

Proof:

Since $\rho_\beta = \eta_\beta = \eta_\gamma = 1$, then $\beta_{00} = \beta_{10} = \beta_{01} = \beta_{11} \triangleq \beta$, $\gamma_{01} = \gamma_{00}$ and $\gamma_{11} = \gamma_{10}$. Thus,

$$VET_{SI} = 1 - \frac{\gamma_{10}(1-\pi) + \gamma_{11}\pi}{\gamma_{00}(1-\pi) + \gamma_{01}\pi} = 1 - \frac{\gamma_{11}}{\gamma_{01}}.$$

So,

$$\begin{aligned} VEA &= 1 - \frac{\Psi_F[\gamma_{10}\alpha_0(1-\pi) + \lambda\gamma_{11}\alpha_1\pi][\beta_{00}(1-\alpha_0)(1-\pi) + \lambda\beta_{01}(1-\alpha_1)\pi]}{[\gamma_{00}(1-\alpha_0)(1-\pi) + \lambda\gamma_{01}(1-\alpha_1)\pi][\beta_{10}\alpha_0(1-\pi) + \lambda\beta_{11}\alpha_1\pi]} \\ &= 1 - \frac{[\gamma_{11}\alpha_0(1-\pi) + \lambda\gamma_{11}\alpha_1\pi][(1-\alpha_0)(1-\pi) + \lambda(1-\alpha_1)\pi]}{[\gamma_{01}(1-\alpha_0)(1-\pi) + \lambda\gamma_{01}(1-\alpha_1)\pi][\alpha_0(1-\pi) + \lambda\alpha_1\pi]} \\ &= 1 - \frac{\gamma_{11}[\alpha_0(1-\pi) + \lambda\alpha_1\pi][(1-\alpha_0)(1-\pi) + \lambda(1-\alpha_1)\pi]}{\gamma_{01}[(1-\alpha_0)(1-\pi) + \lambda(1-\alpha_1)\pi][\alpha_0(1-\pi) + \lambda\alpha_1\pi]} \\ &= 1 - \frac{\gamma_{11}}{\gamma_{01}} = VET_{SI} \quad \text{Q.E.D.} \end{aligned}$$

(6) If $\rho_\beta = 1$ and $\eta_\beta = \eta_\gamma$, then $VE_A = VET_{MAI}$.

Proof:

Since $\eta_\beta = \eta_\gamma$, so $\frac{\beta_{01}}{\beta_{00}} = \frac{\beta_{11}}{\beta_{10}} = \frac{\gamma_{01}}{\gamma_{00}} = \frac{\gamma_{11}}{\gamma_{10}}$. Then we have: $\frac{\gamma_{11}}{\beta_{11}} = \frac{\gamma_{10}}{\beta_{10}} \triangleq a$, $\frac{\gamma_{00}}{\beta_{00}} = \frac{\gamma_{01}}{\beta_{01}} \triangleq b$, and $\frac{\beta_{10}}{\beta_{00}} = \frac{\beta_{11}}{\beta_{01}} = 1$. Then:

$$\begin{aligned} VE_A &= 1 - \frac{\Psi_F[\gamma_{10}\alpha_0(1-\pi) + \lambda\gamma_{11}\alpha_1\pi][\beta_{00}(1-\alpha_0)(1-\pi) + \lambda\beta_{01}(1-\alpha_1)\pi]}{[\gamma_{00}(1-\alpha_0)(1-\pi) + \lambda\gamma_{01}(1-\alpha_1)\pi][\beta_{10}\alpha_0(1-\pi) + \lambda\beta_{11}\alpha_1\pi]} \\ &= 1 - \frac{\Psi_F[a\beta_{10}\alpha_0(1-\pi) + a\lambda\beta_{11}\alpha_1\pi][\beta_{00}(1-\alpha_0)(1-\pi) + \lambda\beta_{01}(1-\alpha_1)\pi]}{[b\beta_{00}(1-\alpha_0)(1-\pi) + b\lambda\beta_{01}(1-\alpha_1)\pi][\beta_{10}\alpha_0(1-\pi) + \lambda\beta_{11}\alpha_1\pi]} \\ &= 1 - \Psi_F \cdot \frac{a}{b} \end{aligned}$$

and,

$$\begin{aligned} VET_{MAI} &= 1 - \frac{\Psi_F[\gamma_{10}(1-\pi) + \lambda\gamma_{11}\pi]}{\gamma_{00}(1-\pi) + \lambda\gamma_{01}\pi} = 1 - \frac{\Psi_F[a\beta_{10}(1-\pi) + a\lambda\beta_{11}\pi]}{b\beta_{00}(1-\pi) + b\lambda\beta_{01}\pi} \\ &= 1 - \Psi_F \cdot \frac{a}{b} \end{aligned}$$

So, $VE_A = VET_{MAI}$.