Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kylie E. C. Ainslie Date

Estimation of influenza vaccine effectiveness from observational studies

By

Kylie E. C. Ainslie

Doctor of Philosophy

Biostatistics

Michael Haber, Ph.D.
Advisor

Vicki Hertzberg, Ph.D.
Committee Member

Andrew Hill, Ph.D.
Committee Member

Limin Peng, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Estimation of influenza vaccine effectiveness from observational studies

By

Kylie E. C. Ainslie
M.S., Emory University, 2016
A.B., Ripon College, 2011

Advisor:
Michael Haber, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2018

Abstract

Estimation of influenza vaccine effectiveness from observational studies
By Kylie E. C. Ainslie

Each year seasonal influenza epidemics cause millions of influenza infections worldwide. Influenza vaccination is recommended as the best way to protect against influenza infection; however, a different vaccine must be produced each season due to changes in influenza virus types, subtypes, and phenotypes from one season to the next. To evaluate the reduction in influenza infections caused by vaccination, vaccine effectiveness (VE) is estimated each year. Observational studies are exclusively used to estimate VE in the United States after the universal recommendation of influenza vaccination because randomized clinical trials are now unethical. Here we consider five different types of observational studies used to assess VE. We use probabilistic models to estimate influenza VE from each study design.

First, we develop a probability model and accompanying maximum likelihood procedure to estimate vaccine-related protection against transmission of influenza from the household and the community. We apply our method to data from a monitored household study conducted in Michigan during the 2012-2013 influenza season and estimate source-specific transmission parameters and VE. We find that the 2012-2013 influenza vaccine provides a significant protective effect against community-acquired transmission.

Second, we develop a dynamic probability model for the evaluation of bias of VE estimates from four commonly used observational study designs: active surveillance cohort (ASC), passive surveillance cohort, test-negative (TN), and traditional case-control. We use the model to evaluate and compare estimates of VE against symptomatic and medically-attended influenza when different sources of bias are present. We show that the preferred study designs for estimating VE against symptomatic influenza and medically-attended influenza are ASC and TN studies, respectively. TN studies are cheaper and involve fewer logistical issues compared to ASC studies; however, if vaccination is suspected to affect the probability of non-influenza acute respiratory illness then one should consider a cohort study.

Finally, we extend our dynamic probability model to further evaluate the bias of VE estimates from test-negative studies. First, we allow vaccination to occur over time (as in a pandemic). Several influenza pandemics have occurred throughout the past century, none greater than in 1918, which killed between 50 and 100 million people. Future influenza pandemics have the potential to inflict a tremendous disease burden. Thus, it is important to determine VE in this setting. Second, we assume an all-or-none vaccine model, where a proportion of vaccinated individuals acquire complete immunity from infection, while the remaining vaccinated individuals acquire no protection. This model differs from the vaccine model assumed in the second topic of this dissertation. However, since vaccine model cannot directly be observed it is important to assess the bias of VE estimates under this alternative model. For each extension we assess the bias of TN-based VE estimates when different sources of bias are present.

Estimation of the effectiveness of influenza vaccination from observational studies

By

Kylie E. C. Ainslie
M.S., Emory University, 2016
A.B., Ripon College, 2011

Advisor:
Michael Haber, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2018

Acknowledgements

Earning a PhD is a solitary pursuit, but one that cannot be accomplished alone. Throughout my long and winding journey to the pinnacle of education, I have been fortunate to have had the unwavering support of a community of people spanning periods in my life, academic disciplines, and geographies. It is not possible to individually thank everyone who has helped me reach the light at the end of, what at times seemed, a very long tunnel. This accomplishment is shared with every person who supported me, believed in me, hugged me, fed and caffeinated me, ran alongside me, talked with me, listened to me, helped me, and loved me.

I would first and foremost like to thank my advisor, Michael Haber, for taking a chance on a late stage graduate student and introducing me to a research area I find fascinating. Thank you for your guidance, advice, time, and support.

Thank you to my committee members Vicki Hertzberg, Andrew Hill, and Limin Peng for their encouragement, input, and assistance.

Thank you to my collaborators, Ryan Malosh, Arnold Monto, Walt Orenstein, and Josh Petrie for endless discussions, edits, and comments, without which this dissertation would not have been possible.

To all of my peers, thank you for enduring the good and bad, difficult and less difficult, failures and successes alongside me. Specifically, thank you to Brooke Alhanti, Nicole Brown, Darcie Cook, Kendra Sirak, Mo Kolawole, and Samantha Noreen for commiserating with me for all these years. My experience at Emory would not have been the same without the friendship of each one of you. Despite the hardships of earning a doctorate, I will look back on my time at Emory fondly because you were a part of this journey.

To all the teachers, coaches, and staff within the City Schools of Decatur, Ripon College, and Emory University who have helped mold and shape me, challenged me, taught me, and supported me, thank you! Thank you to Jeff Shrader for nurturing my love of the intricacies and elegance of mathematics, challenging me to push my mind further, and for showing me how far I can go with hard work. Thank you to Kristen Embry for cultivating my curiosity, challenging my preconceived notions, and for being a friend. Thank you to Amanda Lockhart for making the wonderful world of biology come alive. Thank you to Coach Gary Anderson for recognizing what makes me tick and teaching me the importance of commitment, perseverance, teamwork, and leadership. You are the voice in the back of my head pushing me to "get out of my comfort zone." Thank you to Tim Hess for steering me towards biostatistics despite my initial reticence; Robert Wallace for encouraging me to pursue both my mathematical and biological interests, expecting only the best from me, and increasing my lexicon; and Skip and Michele Wittler for making Ripon, WI my home away from home.

To my family, my parents Vicki and Bill, and my brothers Liam and Griffin, I cannot say thank you enough. I am doubtful that I could have completed this degree without your unconditional love and support. Thank you for enduring me at the worst of times and for celebrating with me at the best of times. I love you.

Thank you to Sarah Jacobs for being there, always.

Finally, my sincerest gratitude goes to Ayanna Abrams for being the calm in a seemingly endless storm, delivering perfect one-liners, and holding up the mirror and turning on the light.

Contents

List of Figures

List of Tables

List of Abbreviations

ACIP	Advisory Committee on Immunization Practices
ARI	acute respiratory illness
ASC	active surveillance cohort
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
FARI	influenza ARI
H1N1	influenza A virus subtype with surface glycoproteins hemagglutinin 1 and neurominidase 1
H3N2	influenza A virus subtype with surface glycoproteins hemagglutinin 3 and neurominidase 2
HA	hemagglitinin
MAI	medically-attended influenza
MH	monitored household
MLE	Maximum likelihood estimate
NA	neurominidase
NFARI	non-influenza ARI
OR	odds ratio
pH1N1	pandemic influenza A (H1N1) strain from 2009 pandemic
PSC	passive surveillance cohort
RCT	randomized clinical trial
RNA	ribonucleic acid
RR	relative risk
RT-PCR	reverse transcription polymerase chain reaction
SI	symptomatic influenza
SE	Standard error
TCC	traditional case-control
TN	test-negative
VE	vaccine effectiveness
WHO	World Health Organization

Chapter 1

Introduction

Each year seasonal influenza epidemics cause millions of influenza infections worldwide [1]. While most individuals who become infected with influenza have mild symptoms and recover without sequelae, some individuals can develop severe illness, particularly those at high risk for infection (e.g., the elderly, persons with chronic health conditions, and young children) [2–5]. An estimated 3 to 5 million cases of severe illness and 300,000 to 650,000 deaths are caused by seasonal influenza epidemics worlwide according to the World Health Organization (WHO) [1, 6]. In the United States, an estimated 140,000 to 710,000 influenza-related hospitalizations and 12,000 to 56,000 influenza-related deaths occur each year [7]. In the 2014-2015 influenza season, a particularly severe influenza season, an estimated 40 million influenza illnesses, 19 million influenza-associated medical visits, and 970,000 influenza-associated hospitalizations occurred in the United States [8].

In addition to seasonal epidemics, influenza has the potential to cause global pandemics. The most recent pandemic in 2009 caused an estimated 60.8 million cases of influenza, 274,304 hospitalizations, and 12,469 deaths in the United States, according to the Centers for Disease Control and Prevention (CDC) [9]. The introduction of a vaccine against the novel influenza virus strain mitigated the pandemic disease burden. An estimated 700,000 to 1.5 million cases, 4,000 to 10,000 hospitalizations, and 200 to 500 deaths were averted by administering

the monovalent (containing a single strain of influenza virus) pandemic vaccine [10].

Due to the rapid genetic changes that influenza viruses undergo, seasonal influenza vaccines must be reformulated and readministered annually. While the effectiveness (i.e., the prevention of illness in vaccinated persons) of influenza vaccines varies from season to season and by individual factors (e.g., age and health status), they are regarded as the most effective way to prevent influenza infection. In 2010, the CDC began recommending annual influenza vaccination to everyone older than six months in the United States [11]. Even in years when influenza vaccine effectiveness (VE) is modest, vaccination can drastically lower the disease burden. For example, the 2015-2016 seasonal influenza vaccine had an effectiveness ranging from 24% to 57% (depending on age group), resulting in an estimated 5.1 million influenza cases, 2.5 million influenza-associated medical visits, and 71,000 hopitalizations that were averted by vaccination [12]. Influenza poses a significant public health burden and vaccination reduces this burden. Thus, it is essential that influenza VE be estimated each season to quantify vaccine performance for the improvement of future vaccines.

1.1 Influenza

Influenza is an infectious disease that can manifest as an acute respiratory illness (ARI) caused by influenza viruses. Influenza viruses are members of the *Orthomyxoviridae* family and have a negative-sense ribonucleic acid (RNA) genome comprised of seven or eight genes [13]. The virus particle is composed of both structural (9) and non-structural (1 or 2) proteins encoded by the genome and

a host-cell derived lipid envelope [14]. Three types of influenza viruses infect humans: A, B, and C; however, infection with influenza C is rare in humans and causes mild symptoms [15]. The two major pathogens in humans are influenza A and B. Influenza A viruses can be further divided into different subtypes based on the composition of glycoproteins, hemagglutinin (HA) and neuraminidase (NA), on the surface of the viral particles. Eighteen HA and 11 NA subtypes have been identified [7]; however, not all combinations are found in humans [16]. Currently, influenza A viruses with surface proteins HA 1 and NA 1 (H1N1) and HA 3 and NA 2 (H3N2) circulate annually [17]. Influenza B viruses are not subtyped, but two distinct lineages of influenza B circulate annually: B/Yamagata and B/Victoria [18].

Influenza viruses circulate annually, typically during the winter months in temperate climates (May to October in the Southern Hemisphere and October to March in the Northern Hemisphere). The seasonality of influenza transmission is dependent on the relative humidity and temperature. Specifically, low relative humidity and cold temperatures (typical of winter months) favor influenza virus transmission because infected individuals shed more virus in colder temperatures and virus particles can remain airborne easier when relative humidity is low [14]. When an infected individual coughs or sneezes, droplets, containing virus particles, are exhaled. When relative humidity is low, water evaporates quickly from exhaled droplets forming very small droplet nuclei, which can remain airborne (and thus be inhaled by nearby individuals). When relative humidity is high, exhaled droplets collect water causing the formation of large droplets which fall more quickly out of the air [19].

In tropical regions, influenza viruses may circulate year-round with some regions experiencing an influenza season that coincides with the rainy season [20,

21], while other locations experience no well-defined influenza season [22]. The different timing of influenza seasons, or lack thereof, in the tropics compared to temperate climates may be caused by different modes of influenza transmission. Transmission in the tropics may be dominated by direct transmission (person-to-person contact), while transmission in temperate regions is dominated by aerosol transmission [23]. It has been shown that aerosol transmission within households is an important mode of transmission in tropical regions [24], so while each region may be dominated by a single mode of transmission, multiple modes play a role in the annual circulation of influenza viruses.

The duration and severity of the influenza season varies from year to year depending on the circulating viral strains [11]. Each year, influenza viruses undergo genetic changes that result in differences in the circulating viruses. Two types of genetic changes can occur, antigenic drift and antigenic shift. Antigenic drift events are small genetic changes that occur continuously as the virus replicates. These small changes produce viruses that are closely related to each other. If a person has immunity against a particular viral strain (either from prior infection or vaccination), and then is later infected with a closely related viral strain, the immune system is more likely to mount a memory response. However, multiple antigenic drift events can accumulate to produce a more distantly related virus that is unlikely to produce a memory response and may make previous immunity from prior exposure or from vaccination ineffective [25].

Antigenic shift events are major genetic changes in the virus resulting in novel viral surface protein combinations (i.e., a novel HA and NA combination), and can only occur in influenza A. An influenza virus that has undergone antigenic shift is very different from the previously circulating viral strains.

Due to the dramatic difference in viral strains, even people who had prior immunity from the original viral strain (before the antigenic shift event) will have little to no immunity or protection against the new strain. Although shift events only occur occasionally [26], these events have the potential to produce viruses that cause large pandemics. For example, an antigenic shift event caused the 2009 H1N1 pandemic influenza strain.

1.1.1 Pandemic Influenza

Since the beginning of the 20th century, there have been five influenza A pandemics: 1918, 1957, 1968, 1977, and 2009. Each pandemic was the result of the emergence, or reemergence, of a novel influenza A strain. The 1918 pandemic (nicknamed the "Spanish" influenza), the most deadly pandemic in history, was caused by the emergence of H1N1 [27, 28]. The 1957 Asian pandemic was caused by the emergence of H2N2, while the 1968 Hong Kong pandemic was characterized by the emergence of H3N2. Usually the emergence of a pandemic strain causes the discontinued circulation of the previously circulating strain. For example, after the 1968 Hong Kong pandemic, H3N2 became the influenza A strain that circulated annually, replacing H2N2. However, in the Russian pandemic of 1977 H1N1 reemerged, and for the first time two influenza A viruses circulated annually (H1N1 and H3N2) [17]. Finally, in 2009 a genetically different pandemic strain of H1N1 (pH1N1) emerged [29].

Influenza pandemics have the potential to cause higher mortality than seasonal epidemics. In the United States, it is estimated that 675,000 people died from the 1918 pandemic, 116,000 from the 1957 pandemic, 100,000 from the 1968 pandemic, and 12,469 people from the 2009 pandemic. The decrease in mortality from each pandemic is likely due, in part, to better disease surveillance,

medical intervention (e.g., vaccines and antivirals), and awareness.

Influenza viruses capable of causing a pandemic can occur in two ways: (1) genetic reassortment between avian influenza viruses and circulating human influenza viruses (as seen in the 1957 and 1968 pandemics) and (2) transmission from an avian reservoir to an intermediate host (such as swine) and then transmission to humans after continued adaptations (as seen in the 1918 pandemic). The 1918 pandemic was so deadly because all eight viral genes came from the avian reservoir, rather than an assortment of avian and human genes. The completely avian genome of the 1918 strain resulted in no immunity in humans [30], hence the catastrophic global mortality (50-100 million people [27]).

Currently, there is no way to prevent the adaptation of influenza viruses and the occurence of another pandemic. The WHO monitors novel influenza A viruses through the Global Influenza Surveillance and Response System that have the potential to adapt the ability to transmit effectively from person-to-person, thus triggering a pandemic. Some novel influenza A viruses are believed to pose a greater threat than others, specifically several avian influenza viruses that have caused infections in humans (H5N1, H7N9, and H9N2) [7]. The WHO and many individual countries, including the United States, have developed pandemic preparedness plans [28]. One element of pandemic preparedness is the rapid development of a vaccine against the pandemic strain. Thousands of hospitalizations and hundreds of deaths were estimated to have been averted by the monovalent vaccine against pH1N1 in 2009 in the United States [10].

1.2 Influenza Vaccine

The CDC recommends the influenza vaccine as the best way to prevent influenza infection [31], and has recommended annual vaccination to everyone over the age of six months in the United States since 2010 [11]. Between the 2005-2006 and 2013-2014 influenza seasons, an estimated 40,127 influenza-associated deaths were averted by influenza vaccination [32]. The influenza vaccine is composed of inactive influenza viruses (both A and B) that are closely related to the strains predicted to circulate in the coming season. Strain predictions in the Northern Hemisphere are made by monitoring the strains that circulate during the winter months in the Southern Hemisphere. Strains are then selected for inclusion in the seasonal vaccine.

1.2.1 History of Influenza Vaccination

The first influenza virus (A(H1N1)) was isolated in 1933. Shortly after the successful isolation of influenza virus, research began on an influenza vaccine. In 1936, the first monovalent influenza vaccine, composed of live-attenuated virus (live virus that has been altered to make the pathogen less virulent or harmless), was created in the Union of Soviet Socialist Republics, and was widely used in factory workers. Due to complexitities in preparing live-attenuated vaccines, efforts were conducted in England and the United States to produce an inactivated vaccine (composed of dead, inactive virus). Following the isolation of influenza B in 1940, the first bivalent influenza vaccine (composed of influenza A and B viruses) was created in 1942 [33]. The first population-scale

use of an inactivated influenza vaccine was in United States military personnel in 1945 [34, 35]. By 1947, the bivalent vaccine began to show limited effectiveness, and a genetically disimilar A(H1N1) virus was isolated, prompting researchers in France to begin studying the vaccine [36]. Due to the genetic variation in influenza viruses from season to season, early influenza vaccines began to show limited protective effects. When H2N2 emerged in 1958, seasonal vaccines again failed, and a new bivalent vaccine was created which contained the new influenza A strain [33].

In 1960, the United States Surgeon General recommended annual influenza vaccination in high-risk individuals following the 1958-1959 influenza pandemic. High-risk individuals were defined as individuals with chronic debilitating diseases, persons over the age of 65, and pregnant women [37]. The Advisory Committee on Immunization Practices (ACIP) reaffirmed this recommendation in 1964 [38]. In 1973, the WHO began issuing recommendations for the composition of influenza vaccines based on influenza surveillance systems.

Due to the co-circulation of two influenza A strains (H1N1 and H3N2) and influenza B in 1978, the first trivalent influenza vaccine was created. Since 2002, two influenza B strains (B/Victoria and B/Yamagata) circulate annually in addition to the two circulating A strains [33]. The ACIP established the first national universal recommendation of seasonal influenza vaccination in 2010 in the United States [11]. In 2012, a quadrivalent vaccine, composed of A(H1N1), A(H3N2), B/Victoria, and B/Yamagata was approved by the United States Food and Drug Administration [33]. Today, both trivalent and quadrivalent vaccines are produced annually.

With the continued production of seasonal influenza vaccines, it is necessary to determine how well each vaccine protects against influenza infection.

The estimation and assessment of influenza VE every season provides information regarding vaccine virus strain similarity with the circulating strains, informs public health officials of the efficacy of the vaccination program in any given season, identifies at risk groups within the population, and provides the basis of educational materials disseminated to the general public and heathcare providers on the merits of seasonal vaccination [39]. Based on the seasonal estimates of VE, the live-attenuated influenza vaccine, previously recommended for children, was not recommended during the 2016-2017 and 2017-2018 seasons due to concerns about its effectiveness against A(pH1N1) during the prior two seasons [40].

1.2.2 Vaccine Efficacy and Effectiveness

Vaccine efficacy is defined as the relative reduction in influenza risk in vaccinated individuals compared to unvaccinated individuals, under ideal conditions [35]. Vaccine *efficacy* refers to the reduction in the risk of disease in a vaccinated person through direct effects of the vaccine assessed through a randomized clinical trial (RCT). Vaccine *effectiveness* refers to the impact of a vaccine assessed using observational studies. Both vaccine efficacy and effectiveness are defined as one minus the relative risk of disease in vaccinated individuals versus unvaccinated individuals. In an outbreak situation, as in a seasonal influenza epidemic, relative risk can be expressed as the ratio of attack rates in the vaccinated and unvaccinated [41]. Attack rate is defined as the proportion of new cases in the susceptible (those capable of being infected) population. Orenstein *et al.* define vaccine efficacy as one minus the ratio of attack rate in vaccinated persons compared to unvaccinated persons [42]. A more appropriate term for the attack rate calculated over the entire influenza season (as is

done when estimating VE) is 'cumulative incidence'. Seasonal attack rates are the cumulative number of infected individuals in a group (e.g., vaccinated or unvaccinated) divided by the total number of persons at risk in that group.

After the univeral recommendation of the influenza vaccine in the Unites States [11], RCTs are no longer considered ethical. Thus, observational studies must be used to estimate vaccine effectiveness rather than efficacy. Both vaccine efficacy and effectiveness measure the direct effects of vaccination (i.e., the amount of protection conferred to an individual by the vaccine). The use of observational studies to assess VE may lead to biased estimates due to lack of random vaccination assignment. However, observational studies allow for the measurement of the effectiveness of a vaccine program through both direct effects of the vaccine and indirect effects on the entire at risk population [43]. The two main classes of observational studies used for the estimation of influeza VE are cohort and case-control studies.

Cohort Studies

In cohort studies used to assess influenza VE, members of the study population are identified prior to the influenza season and followed throughout the study period. There are two main types of cohort studies: active surveillance cohort (ASC) and passive surveillance cohort (PSC). In ASC studies, cases are defined as individuals in the cohort who reported an ARI and test positive for influenza. Study participants report each occurence of ARI symptoms to study personnel and are then tested for influenza infection regardless of whether the participant seeks medical care for their symptoms. In PSC studies, cases are only selected from individuals in the cohort who seek medical care for ARI and subsequently test positive for influenza. Non-cases in both study designs are defined as all

study participants who are not considered cases. ASC studies aim to test everyone who develops ARI for influenza and thus can capture cases with influenza illness of any severity, not just those that necessitate seeking medical care. The prospective nature of cohort studies allows for the collection of information on timing of infection and vaccination.

As in RCTs, VE is defined in cohort studies as one minus the relative risk [44–46], where relative risk is estimated by comparing the incidence of the outcome (e.g., influenza illness) among the vaccinated to the incidence of the outcome among the unvaccinated [47]. VE has also been estimated in cohort studies as one minus the hazard rate ratio [48, 49] using the Cox proportional hazards model [50]; however, the hazard ratio may differ from the relative risk. Unlike in a RCT, vaccination status is not assigned randomly, thus the observed vaccine effect may be confounded by risk factors other than vaccination status [44]. For example, a person's health status may be associated with the probability of being vaccinated. Consider a person who is in good health. This person may be more likely to be vaccinated because they are generally healthy and want to maintain their good health resulting in less vaccinated cases, which bias VE estimates upwards. Conversely, that same healthy person may be less likely to be vaccinated because they are in good health and less likely to become infected with influenza resulting in less unvaccinated cases, biasing VE estimates downwards. The sources of bias that may be present in cohort studies are further discussed in Chapter 3.

Another complication of cohort VE studies is that vaccination status may change during the study period. Estimating relative risk using attack rates makes the implicit assumption that exposure (e.g., vaccination) is the same for each person in the exposed group. However, an individual vaccinated at the

beginning of the study has a different exposure than someone vaccinated in the middle of the study period. An alternative estimate of VE called the person-time estimate (VE_{PT}) allows for vaccination status to vary over time, and is calculated by determining the proportion of new cases in vaccinated and unvaccinated individuals for every time point t. For illustration, we will consider one time point to be one week. Let N_{vj} denote the number of persons in the population of vaccination status $v = 0, 1$ (0 - unvaccinated, 1 - vaccinated) in week $j = 1, \ldots, J$, where J is the last week of the study. Let C_{vj} denote the number of new cases of symptomatic influenza (persons on day 1 of the infectious period) in the population of vaccination status $v = 0, 1$ in week $j = 1, \ldots, J$. Using the above notation, VE can be estimated as

$$VE_{PT} = 1 - \frac{\sum_{j=1}^{J} C_{1j} / \sum_{j=1}^{J} N_{1j}}{\sum_{j=1}^{J} C_{0j} / \sum_{j=1}^{J} N_{0j}}$$

We will use this alternative VE estimate in Chapter 4 to assess the bias of VE estimates when vaccination occurs throughout the study period.

A third type of cohort study, a monitored household (MH) study, is a special type of ASC study in which entire households are enrolled in the study instead of unrelated individuals. A MH study design has recently been adopted to assess influenza VE [49, 51, 52] because transmission of influenza within the household has been shown to play an important role in influenza disease dynamics [24, 53]. During the influenza season, there are potentially two different types of influenza exposures: exposure to other infected household members and exposure to individuals in the larger community. Unlike ASC and PSC

studies, MH studies allow for the separate estimation of VE against household-acquired and community-acquired influenza infection [49]. Determining separate estimates of VE from different sources of infection is the focus of Chapter 2.

Case-Control Studies

Traditional case-control (TCC) studies have historically been used to estimate influenza VE when RCTs were not feasible [54]. Within TCC studies, cases are defined as individuals who seek medical care for an ARI and test positive for influenza infection. Controls are selected randomly from people who did not develop an ARI. In TCC studies, VE is defined as one minus the odds ratio comparing the odds of vaccination in cases and controls [46, 55, 56]. TCC studies may introduce bias into VE estimates because cases are selected only from people infected with influenza who sought medical care, while controls are randomly selected from the population. More recently, a newer case-control study design has been used to estimate influenza VE in an attempt to reduce the bias introduced using a TCC study.

The test-negative design (TN) has become a popular observational study design for estimation of influenza VE [57, 58]. TN studies are a type of case-control study in which cases are defined as individuals who seek medical care for an ARI and test positive for influenza infection. As the name suggests, controls are defined as individuals who seek medical care for ARI and test negative for influenza infection. A TN study was first used to estimate the effectiveness of the pneumococcal vaccine [59], and has since been used to evaluate vaccines against rotavirus [60, 61], cholera [62, 63], meningococcus [64], and influenza

[56, 65]. The TN design was first used to estimate influenza VE in Canada during the 2004-2005 [56] and 2005-2006 [65] influenza seasons. The TN is attractive because it is convenient, can be applied to existing surveillance programs, and controls for confounding due to propensity to seek medical care because cases and controls are selected from the same population. However, selection bias may be introduced into estimates of VE obtained from TN studies because cases and controls are selected only from persons who seek medical care. This may not accurately reflect the entire population of cases and controls.

An important assumption underlying the validity of the TN design is the exposure (e.g., influenza vaccine) has no effect on the probability of developing the control outcome (e.g., a non-influenza ARI (NFARI)). Recent studies to test this assumption are conflicting. In a RCT, Cowling *et al.* found that vaccinees had a significantly increased risk of developing a NFARI compared to non-vaccinees [66]. Other studies found no increased risk of developing a NFARI in vaccinees compared to non-vaccinees [45, 67]. Since the validation of its methodology [55] and core assumption [45], the TN has become the most widely used study design for the estimation of influenza VE.

1.3 Disease Modeling

Before assessing the impact of an intervention on the transmission of an infectious disease, it is important to understand the natural history of the disease (i.e., what causes infection, who is susceptible, and how the disease is transmitted). Influenza infection is caused by influenza viruses. Influenza viruses can infect everyone; however, the young, elderly, and those with chronic health conditions are at increased risk of infection [7]. Influenza is mainly spread

person-to-person by exposure to droplets when infectious individuals cough, sneeze, or talk. Influenza virus particles can survive for hours on non-porous surfaces, so individuals can become infected by touching a surface contaminated with influenza virus and then touching their own mouth or nose [7, 68]. Recent research has shown that aerosol transmission of influenza via very small droplets is an important mode of influenza A transmission, particularly within households [24].

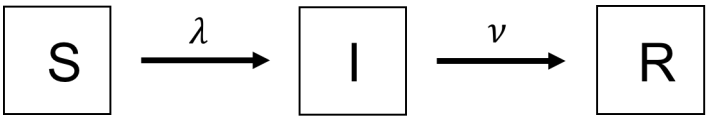
Once an individual becomes infected with influenza, they enter an incubation (or latent) period after which they become infectious. The average latent period for seasonal influenza is two days, but can range from one to four days [69–73]. The infectious period of influenza lasts for approximately four to five days [74], but children or severely immunocompromised individuals may shed virus for longer [75, 76]. After an individual recovers, they usually acquire immunity from the viral strain with which they were infected; however, considerable genetic change in the virus can result in immune escape.

1.3.1 Epidemic Models

Epidemic models are used to approximate the natural history of infectious diseases by combining biological characteristics (e.g., susceptibilitiy, infectiousness, and length of infectious period) with social characteristics (e.g., social mixing patterns and contact structures) [77], and are an important tool used to understand the biological and sociological mechanisms of disease transmission [78]. Many epidemic models divide individuals by their disease status [79–81] and describe the number of individuals who are susceptible to, infected with, and recovered from a particular disease [82]. The simplest epidemic model is

the susceptible-infected-removed (SIR) model, and has been used to model diseases characterized by lifelong immunity after infection, such as measles [83] and whooping cough [84]. The SIR model describes a disease process in which individuals are initially susceptible, can become infected with rate λ, and then after a period of time are removed from the susceptible population either by recovery and acquired immunity or by death with rate ν (Figure 1.1) [79].

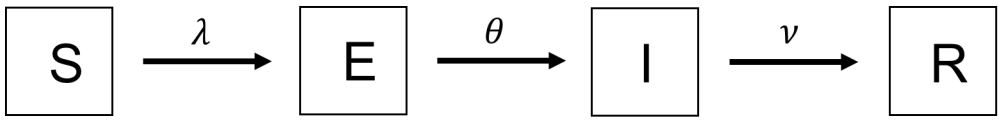
Figure 1.1: SIR epidemic model



SIR model of an infectious disease in which individuals are initially susceptible (S), can become infected (I) with rate λ, and finally are removed (R) from the population with rate ν.

An extension of the SIR model is the susceptible-exposed-infectious-removed (SEIR) model [85], which adds the exposed (E) state, where an individual has been exposed to an infective, but is not yet infectious. In the SEIR model, individuals are initially susceptible and can become infected with rate λ, then after a latent (or exposed) period, may become infectious with rate θ. Finally, infectious individuals recover with rate ν (Figure 1.2). The exposed state is also referred to as the latent period. The SEIR model has been used extensively to model influenza [86–90]. In the following chapters, we assume influenza transmission follows the SEIR model. We build upon this assumption to develop more complex models for the evaluation of vaccine-related protection from influenza infection.

Figure 1.2: SEIR epidemic model



SEIR model of an infectious disease in which individuals are initially susceptible (S), can become exposed (E) with rate λ, can become infectious (I) with rate θ, and finally are removed (R) from the population with rate ν.

1.4 Aims

The aims of this dissertation are to estimate influenza VE from observational studies and assess these estimates with respect to bias. We first estimate influenza VE from different sources of transmission from a MH study (Chapter 2). Next, we assess the bias of VE estimates from four different observational study designs in the presence of sources of bias that may be related to the probabilities of vaccination, developing ARI, and seeking medical care (Chapter 3). We compare the bias from the different study designs to form a recommendation of the study design(s) best used in practice when different sources of bias may be present. Finally, we further investigate the bias of VE estimates from TN studies in Chapter 4 when deviating from our model assumptions in Chapter 3 by allowing vaccination to occur over time and assuming an alternative vaccine model. Throughout this work, we only consider unadjusted estimates of VE as we are interested in characterizing how, and to what degree, sources of bias influence estimates of VE.

1.4.1 Estimation of influenza vaccine effectiveness in household studies

MH studies have gained popularity in the past few influenza seasons [48, 49, 51, 52] since, while logistically more challenging, they provide several advantages over other study designs. MH studies are less susceptible to selection bias (due to enrollment prior to the influenza season), allow for the assessment of VE against influenza infection of any severity, and provide longitudinal data on vaccination status and infection history [49]. The longitudinal nature of household data allows for the estimation of VE against influenza infection from other household members separately from VE against influenza infection from individuals outside of the household.

Previous methods have been developed for the estimation of transmission parameters and VE from household data. Longini and Koopman [91] developed a probability model and maximum likelihood procedure for the separate estimation of influenza transmission parameters in the household and community from final count data, but do not estimate VE. Halloran *et al.* developed a framework to estimate VE from time-to-event household data using the secondary attack rate, but did not account for community transmission [92]. Davis and Haber incorporated temporal information into VE estimates using survival models as a method to estimate influenza transmission probabilities from community and household contacts [93]. Neither of the approaches used by Halloran et al. or Davis and Haber allow for the estimation of source-specific VE, which is the focus of Chapter 2. Ohmit *et al.* [49] estimate source-specific VE using the Cox proportional hazards model, but make assumptions about the source of infection based on viral type/subtype and the timing of infection, as

infection source cannot be directly observed.

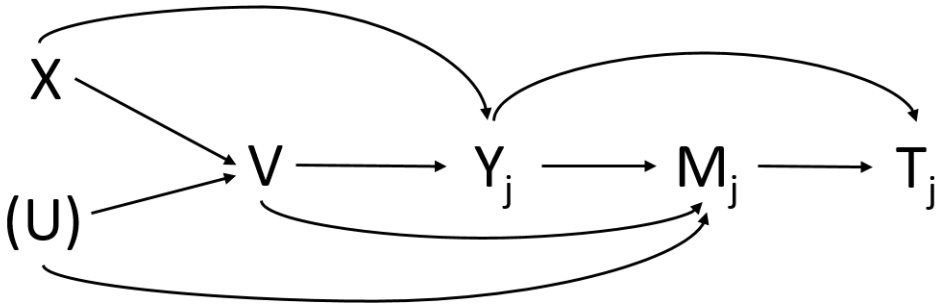
In Chapter 2, we present a probability model and accompanying maximum likelihood procedure to first estimate source-specific transmission parameters from time-to-event household data. Using these transmission parameter estimates, we then estimate VE separately against influenza transmission from the household and from the community. Our approach does not require the source of infection to be known and incorporates temporal information into VE estimates. Additionally, our model allows vaccination to occur during the study. We use symptomatic influenza (SI), defined as laboratory-confirmed infection with the influenza virus that develops into an ARI, as our outcome of interest. Using an agent-based simulation program we evaluate our model. Finally, we apply our method to data from an ongoing, longitudinal MH cohort study in Ann Arbor, MI [49, 51, 52].

1.4.2 Evaluation of bias of influenza vaccine effectiveness estimates from observational studies

Observational studies remain the only option for estimating influenza VE in the United States, most commonly against influenza illness requiring outpatient medical care [94–97], since the CDC's universal recommendation of the seasonal influenza vaccine [11]. Unlike RCTs, the exposure (e.g., vaccination) cannot be randomized in observational studies, possibly introducing bias into estimates of VE. Numerous studies have tried to evaluate the bias of influenza VE estimates from case-control studies [46, 55, 57, 98, 99]. Little has been done to evaluate the bias of VE from cohort studies or to compare the bias of VE estimates in case-control and cohort studies.

In Chapter 3 we evaluate and compare the bias of VE estimates from ASC, PSC, TN, and TCC studies. To accomplish this, we develop a dynamic model (Figure 1.3) that extends previously developed static models [98, 99]. Our model provides several advantages over existing models: 1) a time component allowing for the intensities of FARI and NFARI to change over time and the possibility of developing more than one ARI in a season; 2) the incorporation of two covariates (health status and health awareness) that may affect the probabilities of vaccination, developing ARIs, and seeking medical care for these ARIs; and 3) the ability to assess VE estimates from cohort studies. We use the model to evaluate and compare the bias of VE estimates against both SI and medically-attended influenza (MAI) when different sources of bias are present.

Figure 1.3: Dynamic model of influenza vaccine studies.



X = health status, (U) = health awareness (unobserved), V = vaccination status, Y_j = ARI status in week j, M_j = seeking medical care for ARI in week j, and T_j = influenza test result in week j, where $j = 1, \ldots, J$ and J = the number of weeks in the study.

1.4.3 Bias of test-negative-based estimates of influenza vaccine effectiveness

The TN study design has become the preferred design for assessing annual influenza VE because it is practically easier than other study designs and minimizes confounding related to health-care seeking behavior because both cases and controls are selected from individuals who seek medical care for ARI [100]. Many countries have incorporated TN studies into their annual influenza surveillance systems [101–105]. With the growing popularity of the TN design, it is important to assess the validity of TN-based VE estimates. Previous work has demonstrated the validity of TN-based estimates of VE if vaccination does not affect the probability of developing a NFARI in outpatient settings [45, 55, 57, 98, 106], in inpatient settings [32], and when the influenza test has imperfect sensitivity and specificity [46]. However, these studies make assumptions about the timing of vaccination, the vaccine model, and the definitions of cases and controls. In Chapter 4 we present two extensions to the model presented in Chapter 3.

First, we evaluate the bias of TN-based VE estimates when individuals get vaccinated during the study period, as may happen during a pandemic when a vaccine becomes available in the middle of an outbreak [7]. Previous work addressing the bias of VE estimates from TN studies has assumed that all vaccinated individuals are vaccinated prior to the study period [32, 45, 46, 55, 57, 98, 106]. The assumption of vaccination prior to the study period is reasonable for seasonal epidemics because vaccine campaigns begin prior to the outbreak; however, it is not a reasonable assumption during an influenza pandemic. During the 2009 influenza pandemic, a vaccine was made available months after the

start of the outbreak [7].

Second, we assume an all-or-none vaccine model [44], that is, of the people who are vaccinated a proportion ρ are conferred complete immunity from influenza infection, while the remaining proportion $(1 - \rho)$ are fully susceptible to influenza infection. In Chapter 3 we assumed a leaky vaccine model [44], where vaccinations lowers the probability of infection by ρ.

In Chapter 5 we present on-going work to assess the bias of TN-based VE estimates for different definitions of cases and controls in the presence of sources of bias. Most previous work has ignored different definitions of TN cases and controls. For example, if an individual seeks medical care for ARI twice during the study period and tests negative for influenza during their first visit and positive during their second visit, they could be included only as a control, only as a case, or as a control for their first visit and as a case for their second visit. We aim to assess the bias of TN-based estimates for five different definitions of cases and controls. There is no consensus on the definiton of cases and controls in the event of multiple ARI-related medical visits. This work is motivated by a desire to better inform the design and implementation of TN studies to produce more reliable VE estimates.

Chapter 2

Estimation of influenza vaccine effectiveness in household studies

Originally published in Statistics in Medicine, v. 37(6), 2017, pp. 970-982.
https://doi.org/10.1002/sim.7558

2.1 Introduction

Influenza vaccination is recommended every season due to changes in influenza virus types, subtypes, and phenotypes from one season to the next. The variation in the influenza virus requires the production of a new vaccine each season, thus vaccine effectiveness (VE) must be estimated each year [107]. The concept of 'vaccine effectiveness' is based on comparing the probability of illness of a vaccinated person to that of an unvaccinated person, i.e., it measures the benefit of vaccination for a single individual. VE is defined as one minus the risk ratio, where risk is defined as the probability of becoming infected and ill throughout the influenza season. In this work we use the term 'vaccine effectiveness' rather than 'vaccine efficacy' because the former is estimated using observational studies, while the latter is estimated from a randomized trial.

Observational studies have been increasingly used to assess the benefit of influenza vaccination. Most commonly, observational studies on unrelated individuals have been used to estimate VE against influenza illness requiring outpatient medical care [94–97]. The household unit has been shown to play an important role in the transmission of influenza [24, 53]. Additionally, household data has been shown to provide more robust estimates of VE than data consisting of unrelated individuals [93], thus recent VE studies have employed a MH cohort design [48, 49, 97].

In a MH study, entire households are enrolled and closely monitored over the course of the study period. Whenever a participant has ARI, s/he has to report to study personnel who arrange for a swab to be taken and tested for influenza infection. Despite being expensive and logistically complex, MH studies are attractive to assess influenza VE because they allow for the observation of time of influenza disease and vaccination as well as allow for the estimation of VE against community-acquired and household-acquired influenza [49, 97]. An additional advantage of a MH study is that it allows for the examination of SI of any severity regardless of whether participants sought medical care [49, 94, 97]. Other commonly used study designs (ordinary cohort, in which independent individuals are followed rather than households, case-control, and test-negative [58, 106]) are only able to capture individuals infected with influenza who seek medical care. These studies are prone to bias as many people infected with influenza may not seek medical care, and those who seek medical care might not represent the entire population.

Statistical methods have been developed to estimate influenza VE from household data, first from final data in which influenza infection was assessed after

the season by serological testing [91, 108, 109] and more recently from time-to-event data [49, 51, 92, 93, 97]. Longini and Koopman [91] developed a probability model and maximum likelihood procedure for the separate estimation of influenza transmission parameters in the household and community from final count data. Haber *et al.* [108] extended this model to assess the impact of risk factors on influenza transmission. It has been shown that the use of time-to-event data produces VE estimates with smaller bias compared to estimates produced using final data [93]. Halloran *et al.* [92] developed a framework to estimate VE from time-to-event household data using the secondary attack rate, but did not account for community transmission. Davis and Haber [93] incorporated temporal information into VE estimates using survival models as a method to estimate influenza transmission probabilities from community and household contacts. Neither of these approaches allow for the estimation of source-specific VE, which is the focus of this work. Ohmit *et al.* [49] estimate source-specific VE using the Cox proportional hazards model, but make assumptions about the source of infection based on viral type/subtype and the timing of infection, as infection source cannot be directly observed.

We present a probability model and accompanying maximum likelihood procedure to first estimate source-specific transmission parameters and then to estimate vaccine-related protection against transmission of influenza from the household and from the community from time-to-event household data. Our approach does not require the source of infection to be known and incorporates temporal information into VE estimates. Additionally, our model allows vaccination to occur during the study and does not assume household VE is equal to community VE, thus providing a framework to estimate VE against influenza infection separately from each source. However, for interpretability, VE needs

to be estimated assuming all vaccinated individuals are vaccinated over the entire study period. To assess VE, we use SI, defined as laboratory-confirmed infection with the influenza virus that develops into an ARI, as our outcome of interest. We perform a simulation study to evaluate our model and then apply our model to data from the Household Influenza Vaccine Effectiveness (HIVE) study. The HIVE study has been established in Ann Arbor, MI as an ongoing, longitudinal MH cohort to better characterize the impact of households on influenza transmission [49, 51, 52].

2.2 Methods

We consider a population composed of households of varying sizes as in a MH study. There are potentially two different sources of influenza exposure, exposure to other infected household members and exposure to infected individuals in the larger community. We define the study period as a single influenza season.

We make several important model assumptions: *(1)* Each member of the study population belongs to a household. *(2)* Persons are only classified by their household membership (i.e., there are no other stratifying variables or covariates). *(3)* Each person makes daily contacts with each member of their household and with randomly selected community contacts. *(4)* There is random mixing within the household and among community members. *(5)* A person can only be infected once during the study. Thus, once a person is infected with influenza s/he is removed from the at risk population for the remainder of the study. *(6)* Asymptomatic influenza cases - persons infected with influenza, but do not develop ARI - have a very small probability of transmitting influenza to

others (and therefore, are not accounted for in our model). *(7)* The per-contact transmission probabilities within a household and among the community for vaccinees and non-vaccinees remain constant throughout the study. *(8)* The vaccine provides reduction in transmission probability (leaky vaccine model [43, 44]) and only affects susceptibility to influenza. *(9)* The length of the latent and infectious periods are constant and known.

2.2.1 Probability Model

In real data, it is difficult to ascertain the source of infection for each individual. Recent studies have attributed the source of infection to a household member if the influenza type/subtype are the same and the secondary case was identified within a short time period from the index case [48, 49, 51, 97]. However, it is not possible to actually observe source of infection (except in challenge studies) thus, it is important to develop a probability model that can be used when source of infection is unknown.

Below we present a probability model for estimating VE for household-acquired and community-acquired influenza. To accomplish this, we first estimate influenza transmission parameters within the household and in the community using a maximum likelihood procedure. Then, using these transmission parameter estimates, we estimate source-specific influenza VE. Table 2.1 defines the parameters used in the model.

Consider a susceptible person i on day d. Let Y_{id} denote the infection status for that person, where

Table 2.1: Model parameters

Parameter	Definition
i	Index over people, where $i = 1, \ldots, N$
d	Index over days, where $d = 1, \ldots, D$
j	Index of infection status on a given day: (0=escaped infection (susceptible), 1=made an infectious contact on this day, 2=made an infectious contact before this day).
v_{id}	Vaccination status of person i on day d. Also denoted at v (0=unvaccinated, 1=vaccinated)
β_v	Daily transmission probability from an infectious household member to a susceptible with vaccination status $v = 0, 1$
γ_v	Daily transmission probability from infectious community contacts to a susceptible with vaccination status $v = 0, 1$
m_{id}	The number of infectious persons in the household of person i on day d
$p(d)$	The prevalence of influenza infection on day d
π_{ijd}	Conditional probability that person i has infection status j on day d given that s/he was susceptible on the previous day
ψ_{ijd}	Unconditional probability that person i has infection status j on day d
λ_{iH}	Probability that person i was infected from a household contact by the end of the study
λ_{iC}	Probability that person i was infected from a community contact by the end of the study

$$Y_{id} = \begin{cases} 0 & \text{person } i \text{ susceptible by the end of day } d \\ 1 & \text{person } i \text{ made an infectious contact on day } d \\ 2 & \text{person } i \text{ made an infectious contact before day } d \end{cases}$$

Let v_{id} denote her/his vaccination status ($v_{id} = 0, 1$ for unvaccinated and vaccinated, respectively). We define $\beta_{v_{id}}$ as the daily transmission probability to that person from a single household contact; similarly, we define $\gamma_{v_{id}}$ as the

daily transmission probability of infection from the community to that person when everyone else is infectious. Since v_{id} can only take on the values $0, 1$ for any person on any day, we have four transmission parameters: $\beta_0, \beta_1, \gamma_0, \gamma_1$.

We assume that the latent period, the time between an individual getting infected and becoming infectious, begins on the day after the infectious contact was made. An infected person becomes *infectious* $L + 1$ days after making an infectious contact and remains infectious for I days, where L and I are the length of the latent and infectious periods, respectively. For example, if an infectious contact is made on day 1 then, the latent period lasts two days (days 2 and 3) and the infectious period lasts four days (days 4-7). After the duration of the infectious period, the person recovers and remains immune for the rest of the study. It is usually assumed that the day of becoming infectious is the day of onset of symptoms, i.e., the length of the incubation period equals the length of the latent period. During our estimation process, we observe the first day of the infectious period and determine the day of the infectious contact by subtracting $L + 1$. We let $p(d)$ denote the prevalence of influenza infection on day d, defined as the proportion of the population who is infectious [110], and let m_{id} be the number of infectious persons in the household of person i on day d. When determining the probability that a person is infected from community contacts on day d, we multiply $\gamma_{v_{id}}$ by the proportion of infectious individuals in the population on that day, $p(d)$.

Let $\pi_{ijd} = \mathbb{P}(Y_{id} = j | Y_{i(d-1)} = 0)$ denote the conditional probability that person i has infection status j on day d, given that s/he was susceptible on the previous day, $j = 0, 1, 2$. Let $\psi_{ijd} = \mathbb{P}(Y_{id} = j)$ denote the unconditional probability that person i has infection status j on day d, $j = 0, 1, 2$. All of the probabilities involving individual persons are conditioned on the vaccination

history of individual i, $\mathbf{V}_i = (V_{i1}, \ldots, V_{iD})$, $i = 1, \ldots, N$.

The conditional probabilities π_{ijd} can be written as follows:

$$\begin{aligned}
\pi_{i0d} &= \mathbb{P}(Y_{id} = 0 | Y_{i(d-1)} = 0) &= (1 - \beta_{v_{id}})^{m_{id}}(1 - \gamma_{v_{id}} \cdot p(d)) \\
\pi_{i1d} &= \mathbb{P}(Y_{id} = 1 | Y_{i(d-1)} = 0) &= 1 - [(1 - \beta_{v_{id}})^{m_{id}}(1 - \gamma_{v_{id}} \cdot p(d))] \\
\pi_{i2d} &= \mathbb{P}(Y_{id} = 2 | Y_{i(d-1)} = 0) &= 0
\end{aligned}$$

where, given person i was susceptible on day $d-1$, π_{i0d} is the probability of person i escaping infection on day d, π_{i1d} is the probability of person i becoming infected on day d, and π_{i2d} is the probability that person i was infected on a previous day. Under assumption *(5)* $\pi_{i2d} = 0$.

The unconditional probability of person i having infection status j ($j = 0, 1$) on day d is defined as

$$\begin{aligned}
\psi_{ijd} &= \mathbb{P}(Y_{id} = j) \\
&= \mathbb{P}(Y_{id} = j | Y_{i(d-1)} = 0)\mathbb{P}(Y_{i(d-1)} = 0) + \mathbb{P}(Y_{id} = j | Y_{i(d-1)} > 0)\mathbb{P}(Y_{i(d-1)} > 0).
\end{aligned}$$

By assumption (5), $\mathbb{P}(Y_{id} = j | Y_{i(d-1)} > 0) = 0$, thus

$$\begin{aligned}
\mathbb{P}(Y_{id} = j) &= \mathbb{P}(Y_{id} = j | Y_{i(d-1)} = 0)\mathbb{P}(Y_{i(d-1)} = 0) \\
&= \pi_{ijd}\psi_{i0(d-1)}
\end{aligned}$$

For example, the probability that person i, who was effectively vaccinated on day 2, gets infected on day 3 is

$$\begin{aligned}
\mathbb{P}(Y_{i3} = 1) &= \pi_{i13}\psi_{i02} \\
&= \pi_{i13}\pi_{i02}\pi_{i01} \\
&= \{1 - [(1 - \beta_1)^{m_{i3}}(1 - \gamma_1 p(3))]\}\{(1 - \beta_1)^{m_{i2}}(1 - \gamma_1 p(2))\}\{(1 - \beta_0)^{m_{i1}}(1 - \gamma_0 p(1))\},
\end{aligned}$$

where the value of v_{id} changes from 0 to 1 on day 2.

2.2.2 Maximum Likelihood Procedure

Each person's contribution to the likelihood function depends on whether or not s/he became infected during the study and on the day of infection, if infected. If person i got infected on day d then his/her contribution to the likelihood function is: $L_i = \psi_{i1d}$. If person i did not become infected by the last day of the study D, $L_i = \psi_{i0D}$, the probability of escaping infection throughout the study. The overall likelihood is $L(\beta_0, \beta_1, \gamma_0, \gamma_1|\text{data}) = \prod_{i=1}^{N} L_i$, where N is the number of study participants. We assume that persons are (conditionally) independent because our probabilities condition on the daily number of infected persons in the household and the daily prevalences of infection in the community. Finally, maximum likelihood estimates (MLEs) of the transmission parameters, $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_0, \hat{\gamma}_1$, are obtained by maximizing $L(\beta_0, \beta_1, \gamma_0, \gamma_1|\text{data})$. Likelihood optimization was performed using a limited-memory modification of the BFGS quasi-Newton method [111] with a lower bound of 0 and an upper bound of 1 using the R function optim [112]. Standard errors (SEs) of transmission parameter estimates were obtained from the Hessian matrix from the maximization procedure, empirically from simulations by taking the standard deviation of all simulation estimates, and by parametric bootstrap. Using the transmission parameter MLEs, we can estimate the distribution of Y_{id} for every (i, d) by plugging the parameter estimates into the equations for π_{ijd} and ψ_{ijd}.

We estimate VE using a two-step process. First, we estimate the transmission probabilities from the likelihood function in which a person's actual day of vaccination, before or during the study, is used. Second, using the estimated

transmission probabilities from Step 1, we estimate VE by comparing the probability of becoming infected during the entire study between persons who became effectively vaccinated prior to the study and persons who remained unvaccinated throughout the study. In this way, the estimates of VE do not depend on the times of vaccination. For each person i, we added to the population two dummy persons: person A_i who was effectively vaccinated before the onset of the study, and person B_i who remained unvaccinated throughout the study. These dummy persons make the same contacts with real persons (but not with other dummy persons) that correspond with the contacts of person i. They can become infected but they are unable to infect others. Therefore, the dummy persons do not affect the infection probabilities of the real persons (i.e., they do not modify the vaccine's indirect effects). The vaccination status of all real persons remained unchanged for the purpose of estimating VE. We define λ_{iH} and λ_{iC} as the probability that person i is infected from a household (H) or community (C) contact during the study, respectively. The MLEs of λ_{iH} and λ_{iC} are obtained by substituting the MLEs of our transmission parameters for the true parameters (see Appendix A for details). Due to the very small probability of coinfection we do not include the probability of being coinfected in estimates of VE.

We denote the seasonal VEs against household transmission, community transmission, and overall transmission by VE_H, VE_C, and VE_O, respectively.

The estimates of VE are

$$\hat{VE}_H = 1 - \frac{\sum_{i=1}^{N} \hat{\lambda}_{A_i H}}{\sum_{i=1}^{N} \hat{\lambda}_{B_i H}}$$

$$\hat{VE}_C = 1 - \frac{\sum_{i=1}^{N} \hat{\lambda}_{A_i C}}{\sum_{i=1}^{N} \hat{\lambda}_{A_i C}}$$

$$\hat{VE}_O = 1 - \frac{\sum_{i=1}^{N} (\hat{\lambda}_{A_i H} + \hat{\lambda}_{A_i C})}{\sum_{i=1}^{N} (\hat{\lambda}_{B_i H} + \hat{\lambda}_{B_i C})},$$

SEs for VE estimates were obtained empirically and by parametric bootstrap.

2.2.3 Simulations

To assess the performance of our method and the accuracy of our maximum likelihood estimates, we developed a stochastic agent-based simulation program to simulate an influenza outbreak in a population with households. One simulation corresponded to one outbreak. Each simulation featured a susceptible population of 1000 individuals with 10 initially infected individuals. Each individual was assigned to a household. Households varied in size from one to twelve members. The proportion of households of each size were based on DeKalb County, GA census data [113]. The influenza season lasted three months. It was assumed that each person made 10 daily community contacts (under the assumption of random-mixing) and made daily contact with each

person in their household. Two vaccination scenarios were assessed: (1) all vaccinations occurred prior to the study period and (2) vaccinations occurred prior to the study (25%), during the first month (15%), and during the second month (10%). In each vaccination scenario, 50% of the population was vaccinated. A person was considered effectively vaccinated 14 days after the receipt of the influenza vaccine. The following parameter values were used as the daily transmission probability from an infectious person to a susceptible person with vaccination status v (v=0, unvaccinated, v=1, vaccinated): $\alpha_0 = 0.04$ ($\gamma_0 = \alpha_0 \cdot 10 = 0.40$), $\alpha_1 = 0.01$ ($\gamma_1 = \alpha_1 \cdot 10 = 0.1$), $\beta_0 = 0.15$, and $\beta_1 = 0.075$.

MLEs and SEs of the transmission parameters were calculated for each simulation. The MLEs of the transmission parameters were used to estimate household VE, community VE, and overall VE. True VE was calculated using the true transmission parameter values under each vaccination scenario. For each simulation scheme, the bias of the VE estimates was assessed. Bias was defined as the true VE subtracted from the estimated VE. Two hundred outbreak simulations were performed and fifty bootstrap simulations were performed for each outbreak simulation. Source specific VE, SE, and 95% confidence intervals (CI) were estimated for each simulation. The same assumptions we made for our model (see section 2) were used for the simulation program. The latent period was set to 2 days and the infectious period was set to 4 days [74].

2.2.4 Sensitivity Analyses

Sensitivity analyses were performed to assess the bias of VE estimates using the maximum likelihood procedure when several model assumptions were relaxed. Each sensitivity analysis was conducted independently. In the first sensitivity analysis, we relaxed assumption *(10)* (the latent and infectious periods

fixed and known) and allowed the latent and infectious periods to follow a distribution. We assessed bias under two different scenarios: (1) the mean latent and infectious periods were correctly specified and (2) the mean latent and infectious periods were incorrectly specified. For scenario (1) the latent period was 1, 2, or 3 days with probabilities 0.25, 0.5, and 0.25, respectively. The infectious period was 3, 4, 5, or 6 days with probabilities 0.3, 0.5, 0.1, and 0.1, respectively. The mean latent and infectious periods were 2 and 4 days, respectively. For scenario (2) the latent period was 1, 2, or 3 days with probabilities 0.1, 0.1, and 0.8, respectively with a mean of 2.7 days. The infectious period was 3, 4, 5, or 6 days with probabilities 0.05, 0.1, 0.65, and 0.2, respectively with a mean of 5 days. The specified mean latent and infectious periods, used for the derivation of the MLEs, were 2 and 4 days, respectively.

In the second sensitivity analysis, we assessed the bias of VE estimates when the prevalence in the cohort was allowed to differ from the prevalence in the overall population. An overall population of households, comprised of 2000 individuals, was simulated, from which households were selected to make up the study cohort of 1000 individuals. Households in the cohort were assumed to be a random sample of the households in the overall population. MLEs, SEs, and 95% CIs were calculated for each simulation.

2.2.5 A Real-Life Example

Our maximum likelihood approach was applied to a dataset from the HIVE study in Michigan during the 2012-2013 influenza season designed to estimate household and community VE. The study population consisted of 321 households with a total sample size of 1426 members followed from October 2012 to May 2013. Households ranged in size from 4 to 10 members. Only households

with at least 4 persons including at least 2 children were included in the study. At the onset of influenza-like symptoms, participants were instructed to contact study personnel, so a respiratory specimen could be collected. Specimens were tested for the presence of influenza virus by reverse transcription polymerase chain reaction (RT-PCR) [49, 51, 52, 97]. Influenza infection was identified by RT-PCR in 111 individuals with influenza-like illness. Five individuals were infected with influenza twice [49], but only the first influenza infection was considered for our analysis. Index cases were assumed to be infected from the community and a household-acquired case was defined by transmission link to an index case within the household if both cases were caused by the same influenza type/subtype/lineage and if illness onset in the secondary case occurred 1-7 days after illness onset in the index case. Vaccination status was determined as previously described using a combination of medical records and state registry data. Adults and children aged 9-17 years old were considered effectively vaccinated 14 days after the receipt of the vaccine. Children under the age of 9 years old were considered effectively vaccinated 14 days after receipt of the second dose of the vaccine [49].

Influenza transmission parameters for vaccinated and unvaccinated individuals and VE against household-acquired and community-acquired transmission of influenza were estimated using our maximum likelihood approach. To avoid undefined values during maximum likelihood estimation due to a prevalence of zero within the cohort, values of zero were changed to $1/1426$ (the total size of the cohort). Fifty parametric bootstrap simulations were performed to obtain SE estimates and 95% CIs for transmission parameter and VE estimates. Simulations were performed using information from the study data, such as, transmission parameter estimates, proportion of households of each

size and proportion of individuals vaccinated. Model adequacy was assessed by comparing the mean number of cases per household size from 200 simulated outbreaks to the observed frequencies of cases per household size in the data. All simulations and analyses were performed in R 3.2.2 [112].

2.3 Results

2.3.1 Simulations

Mean transmission parameter and SE estimates from 200 simulations are shown in Table 2.2. When all vaccinations occurred prior to the study, our maximum likelihood procedure produced the following estimates: $\hat{\beta}_0 = 0.153$ (95% CI: 0.129, 0.177), $\hat{\beta}_1 = 0.078$ (95% CI: 0.064, 0.092), $\hat{\gamma}_0 = 0.429$ (95% CI: 0.380, 0.478), $\hat{\gamma}_1 = 0.118$ (95% CI: 0.092, 0.142). When vaccinations occurred prior to and during the study, $\hat{\beta}_0 = 0.156$ (95% CI: 0.132, 0.180), $\hat{\beta}_1 = 0.078$ (95% CI: 0.060, 0.096), $\hat{\gamma}_0 = 0.425$ (95% CI: 0.382, 0.468), $\hat{\gamma}_1 = 0.118$ (95% CI: 0.093, 0.143). Transmission parameter estimates were similar between the two vaccination scenarios. All transmission parameter estimates were close to the true values ($\beta_0 = 0.15$, $\beta_1 = 0.075$, $\gamma_0 = 0.4$, $\gamma_1 = 0.10$). The greatest bias observed was in the estimate of γ_0 when vaccinations occurred prior to and during the study (Bias=0.031 corresponding to a relative bias of 7.75%); however, the estimates of γ_1 suffered from relative biases of 16% and 17% in vaccination scenarios (1) and (2), respectively (Table 2.2). SE estimates were calculated empirically, using the Hessian matrix from the maximum likelihood procedure, and from fifty parametric bootstraps. All three SE estimation methods produced similarly small SE estimates (Table 2.2).

Table 2.2: Household and community transmission parameter estimates by vaccination status from 200 simulated influenza outbreaks

Vaccination Scenario	Value	β_0	β_1	γ_0	γ_1
	True	*0.150*	*0.075*	*0.400*	*0.100*
	Estimate	*0.153*	*0.078*	*0.429*	*0.118*
	(95% CI)	(0.129, 0.177)	(0.064, 0.092)	(0.380, 0.478)	(0.092, 0.142)
Before study	Bias	0.003	0.003	0.029	0.018
	SE (Empirical)	0.012	0.007	0.025	0.011
	SE (Hessian)	0.013	0.007	0.026	0.011
	SE (Bootstrap)	0.014	0.007	0.027	0.012
	Estimate	*0.156*	*0.078*	*0.425*	*0.118*
Before and during study	(95% CI)	(0.132, 0.180)	(0.060, 0.096)	(0.382, 0.468)	(0.093, 0.143)
	Bias	0.006	0.003	0.025	0.018
	SE (Empirical)	0.012	0.009	0.022	0.013
	SE (Hessian)	0.012	0.008	0.023	0.014
	SE (Bootstrap)	0.012	0.009	0.022	0.015

Transmission parameters (β_0-household, unvaccinated; β_1-household, vaccinated; γ_0-community, unvaccinated; and γ_1-community, vaccinated), 95% CIs, and SEs. SEs were calculated empirically, using the Hessian matrix from the maximization procedure, and by parametric bootstrap.

Mean VE and SE estimates from 200 simulations are shown in Table 2.3. When vaccinations occurred prior to the study, our maximum likelihood procedure produced the following estimates: $\hat{VE}_{MLH} = 0.478$ (95% CI: 0.358, 0.598), $\hat{VE}_{MLC} = 0.724$ (95% CI: 0.659, 0.789), and $\hat{VE}_{MLO} = 0.607$ (95% CI: 0.546, 0.668). When vaccinations occurred prior to and during the study, $\hat{VE}_{MLH} = 0.453$ (95% CI: 0.314, 0.592), $\hat{VE}_{MLC} = 0.719$ (95% CI: 0.652, 0.786), and $\hat{VE}_{MLO} = 0.602$ (95% CI: 0.535, 0.669). The greatest bias was observed in the estimate of VE against community-acquired influenza when all vaccinations occurred prior to the study, Bias=-0.024 (corresponding to a relative bias of 3.2%). SEs of the

VE estimates were calculated empirically and via parametric bootstrap. Empirical SEs were very similar to bootstrap SEs. VE against household-acquired influenza consistently had the highest SE, while VE estimates against community-acquired and overall influenza were very similar. When vaccinations occurred during the study, the empirical SEs were slightly larger for all VE estimates than when all vaccinations occurred prior to the study. For all estimates of VE, we observed coverage probabilities of or greater than 95% with the exception of community VE when all vaccinations occurred prior to the study (Table 2.3).

Table 2.3: Maximum likelihood VE estimates against influenza infection in the household, community, and overall from 200 simulated influenza outbreaks

Vaccination Scenario	**Value**	$\mathbf{VE_{MLH}}$	$\mathbf{VE_{MLC}}$	$\mathbf{VE_{MLO}}$
Before study	*True*	*0.477*	*0.746*	*0.607*
	Estimate	*0.467*	*0.721*	*0.601*
	(95% CI)	(0.354, 0.580)	(0.662, 0.779)	(0.544, 0.659)
	Bias	-0.010	-0.025	-0.006
	SE (Empirical)	0.056	0.030	0.030
	SE (Bootstrap)	0.064	0.034	0.030
	Coverage Probability	0.960	0.920	0.960
Before and during study	*True*	*0.470*	*0.745*	*0.605*
	Estimate	*0.469*	*0.717*	*0.602*
	(95% CI)	(0.339, 0.600)	(0.645, 0.789)	(0.540, 0.664)
	Bias	-0.001	-0.028	-0.003
	SE (Empirical)	0.067	0.037	0.032
	SE (Bootstrap)	0.069	0.038	0.033
	Coverage Probability	0.960	0.950	0.970

Household VE (VE_{MLH}), community VE (VE_{MLC}), overall VE (VE_{MLO}), 95% CIs, and SEs. SEs were calculated empirically and by parametric bootstrap.

2.3.2 Sensitivity Analyses

Table 2.4 shows the results of the sensitivity analyses performed when the assumption that the latent and infectious periods are fixed and known was relaxed. When the mean lengths of the latent and infectious periods were correctly specified and all vaccinations occurred prior to the study, $\hat{VE}_{MLH} = 0.474$ (95% CI: 0.348, 0.599), $\hat{VE}_{MLC} = 0.703$ (95% CI: 0.635, 0.770), and $\hat{VE}_{MLO} = 0.602$ (95% CI: 0.546, 0.657). When vaccinations occurred prior to and during the study, $\hat{VE}_{MLH} = 0.470$ (95% CI: 0.327, 0.612), $\hat{VE}_{MLC} = 0.699$ (95% CI: 0.630, 0.768), and $\hat{VE}_{MLO} = 0.600$ (95% CI: 0.531, 0.669). When the mean lengths of the latent and infectious periods were misspecified and all vaccinations occurred prior to the study, $\hat{VE}_{MLH} = 0.440$ (95% CI: 0.303, 0.577), $\hat{VE}_{MLC} = 0.679$ (95% CI: 0.617, 0.740), and $\hat{VE}_{MLO} = 0.589$ (95% CI: 0.541, 0.636). When vaccinations occurred prior to and during the study, $\hat{VE}_{MLH} = 0.442$ (95% CI: 0.311, 0.573), $\hat{VE}_{MLC} = 0.657$ (95% CI: 0.588, 0.726), and $\hat{VE}_{MLO} = 0.575$ (95% CI: 0.524, 0.626) (Table 2.4).

When the latent and infectious periods were allowed to follow a distribution, VE estimates were underestimated. When the mean latent and infectious periods were correctly specified the largest bias was observed in estimates of VE against community transmission with a relative bias of 6%. Larger bias was observed in all VE estimates when the mean lengths of the latent and infectious periods were misspecified with the largest bias (corresponding to a relative bias >11%) observed in estimates of VE against community transmission when vaccination occurred prior to and during the study. Estimates of SE were similar when the mean latent and infectious periods were correctly specified and misspecified. Estimates of VE_{MLH} had the largest SE and the SE of VE_{MLC} and VE_{MLO} were similar (Table 2.4).

Table 2.4: Estimates of VE when the latent and infectious periods are not constant from 200 simulated influenza outbreaks

Mean Lengths of Latent and Infectious Periods	Vaccination Scenario	Value	VE_{MLH}	VE_{MLC}	VE_{MLO}
$\bar{L} = 2, \bar{I} = 4$	Before study	*True*	*0.477*	*0.747*	*0.607*
		Estimate	*0.474*	*0.703*	*0.602*
		(95% CI)	(0.349, 0.599)	(0.635, 0.770)	(0.546, 0.657)
		Bias	-0.003	-0.044	-0.005
		SE (Empirical)	0.064	0.035	0.028
		SE (Bootstrap)	0.074	0.031	0.032
	Before and during study	*True*	*0.471*	*0.745*	*0.606*
		Estimate	*0.470*	*0.699*	*0.600*
		(95% CI)	(0.327, 0.612)	(0.630, 0.768)	(0.531, 0.669)
		Bias	-0.001	-0.046	-0.006
		SE (Empirical)	0.073	0.035	0.035
		SE (Bootstrap)	0.070	0.034	0.033
$\bar{L} = 2.7, \bar{I} = 5$	Before study	*True*	*0.469*	*0.745*	*0.605*
		Estimate	*0.440*	*0.679*	*0.589*
		(95% CI)	(0.303, 0.577)	(0.617, 0.740)	(0.542, 0.636)
		Bias	-0.029	-0.066	-0.016
		SE (Empirical)	0.070	0.031	0.024
		SE (Bootstrap)	0.074	0.028	0.027
	Before and during study	*True*	*0.462*	*0.743*	*0.604*
		Estimate	*0.442*	*0.657*	*0.575*
		(95% CI)	(0.298, 0.559)	(0.584, 0.720)	(0.525, 0.623)
		Bias	-0.020	-0.086	-0.029
		SE (Empirical)	0.067	0.035	0.026
		SE (Bootstrap)	0.075	0.033	0.030

Bias of VE estimates were calculated allowing the latent and infectious periods to follow a distribution. VE_{MLH}, VE_{MLC}, and VE_{MLO} denote VE estimates using the maximum likelihood approach against household, community, and overall transmission, respectively. We considered two situations. First, the mean length of the latent ($\bar{L}$) and infectious ($\bar{I}$) periods were $\bar{L} = 2$ and $\bar{I} = 4$ (i.e., they were correctly specified). Second, $\bar{L} = 2.7$ and $\bar{I} = 5$ (i.e., they were misspecified). In both situations, during the estimation procedure, it was assumed that L=2 and I=4. VE estimates, 95% CIs, and SEs were obtained from 200 simulations under two different vaccination scenarios. SEs were calculated empirically and by parametric bootstrap.

Table 2.5 shows the results of the sensitivity analysis when the cohort was selected from a simulated overall population. This simulation scenario allowed

the prevalence of influenza infection in the cohort to differ from the prevalence in the overall population. When all vaccinations occurred prior to the study, $\hat{VE}_{MLH} = 0.480$ (95% CI: 0.344, 0.615), $\hat{VE}_{MLC} = 0.727$ (95% CI: 0.662, 0.791), and $\hat{VE}_{MLO} = 0.614$ (95% CI: 0.551, 0.677). When vaccinations occurred prior to and during the study, $\hat{VE}_{MLH} = 0.469$ (95% CI: 0.334, 0.604), $\hat{VE}_{MLC} = 0.722$ (95% CI: 0.648, 0.796), and $\hat{VE}_{MLO} = 0.607$ (95% CI: 0.544, 0.671) (Table 2.5). The biases of VE estimates when the cohort was selected from a larger overall population were very similar to the original simulations in which only the cohort was simulated (Tables 2.5 and 2.3, respectively).

Table 2.5: VE estimates from a random sample drawn from a larger population from 200 simulated influenza outbreaks

Vaccination Scenario	Value	$\mathbf{VE_{MLH}}$	$\mathbf{VE_{MLC}}$	$\mathbf{VE_{MLO}}$
Before study	*True*	*0.476*	*0.746*	*0.605*
	Estimate	*0.480*	*0.727*	*0.614*
	(95% CI)	(0.344, 0.615)	(0.662, 0.791)	(0.551, 0.677)
	Bias	0.004	-0.019	0.009
	SE (Empirical)	0.069	0.033	0.032
	SE (Bootstrap)	0.065	0.032	0.030
Before and during study	*True*	*0.469*	*0.745*	*0.604*
	Estimate	*0.469*	*0.722*	*0.607*
	(95% CI)	(0.334, 0.604)	(0.648, 0.796)	(0.544, 0.671)
	Bias	0.000	-0.023	0.003
	SE (Empirical)	0.069	0.038	0.032
	SE (Bootstrap)	0.070	0.037	0.032

Bias of VE from a sample population randomly selected from a larger population. VE_{MLH}, VE_{MLC}, and VE_{MLO} denote VE estimates using the maximum likelihood approach against household, community, and overall transmission, respectively. True VE was calculated from the overall population. VE estimates, 95% CIs, and SEs were obtained from 200 simulations performed under two different vaccination scenarios. SEs were calculated empirically and by parametric bootstrap.

2.3.3 A Real-Life Example

Estimates of transmission parameters from the HIVE study are shown in Table 2.6. The daily transmission probability from an infectious household contact to an unvaccinated susceptible is 0.013 (95% CI: 0.008, 0.019) and to a vaccinated susceptible is 0.013 (95% CI: 0.005, 0.021). The transmission rate from all daily infectious community contacts to an unvaccinated susceptible is 0.202 (95% CI: 0.159, 0.245) and to a vaccinated susceptible is 0.134 (95% CI: 0.081, 0.187) (Table 2.6). A transmission rate of 0.202 (0.134) means that, on average, 20.2% (13.4%) of unvaccinated (vaccinated) persons who make contacts with infectious persons on a given day will become infected. The lower estimated transmission rate in the community to vaccinated individuals compared to unvaccinated individuals suggests that there is at least a small protective effect of vaccination against community-acquired influenza infection. SE estimates were obtained using the Hessian matrix from the maximum likelihood procedure and using parametric bootstrap. The bootstrap SE estimates were similar to the Hessian matrix SE estimates (Table 2.6). Our simulated frequencies of cases per household size (Table A.1) were a good match to the observed frequencies (Table A.2) suggesting that our model captures the dependency between household size and attack rate.

Estimates of household VE, community VE, and overall VE are presented in Table 2.7. VE point estimates indicated significant protection against community-acquired influenza infection (0.336, 95% CI: 0.066, 0.606), and non-significant protection against household-acquired (0.052, 95% CI: -0.754 0.858) and overall (0.250, 95% CI: -0.019, 0.519) influenza infection (Table 2.7).

Table 2.6: Transmission parameter estimates from HIVE Study data

Value	β_0	β_1	γ_0	γ_1
Estimate	0.013	0.013	0.202	0.143
(95% CI)	(0.008, 0.019)	(0.005, 0.021)	(0.159, 0.245)	(0.081, 0.187)
SE (Hessian)	0.003	0.003	0.032	0.026
SE (Bootstrap)	0.003	0.004	0.022	0.027

Maximum likelihood transmission parameter estimates, 95% CIs, and SEs from HIVE Study data (2012-2013). Transmission parameters were defined as follows: β_0-household, unvaccinated; β_1-household, vaccinated; γ_0-community, unvaccinated; and γ_1-community, vaccinated.

Table 2.7: Maximum likelihood estimates of VE from HIVE Study data

Value	$\mathbf{VE_{MLH}}$	$\mathbf{VE_{MLC}}$	$\mathbf{VE_{MLO}}$
Estimate	0.052	0.336	0.250
(95% CI)	(-0.754, 0.858)	(0.066, 0.606)	(-0.019, 0.519)
SE	0.411	0.138	0.137

Maximum likelihood VE estimates, 95% CIs, and SEs from HIVE Study data (2012-2013) [49]. VE_{MLH}, VE_{MLC}, and VE_{MLO} denote VE estimates using the maximum likelihood approach against household, community, and overall transmission, respectively.

We compared our VE estimates to results found by Ohmit *et al.* [49] using unadjusted and adjusted hazard rate ratios (Table 2.8). The adjusted models adjusted for age in months and documentation of high-risk health status [49]. Our point estimate of VE against household-acquired influenza was substantially lower (by more than 0.25) than the estimates found by Ohmit *et al.*. Our point estimate of VE against community-acquired influenza infection was slightly higher than the estimates in the original paper, and we were able to

detect a significant protective effect of vaccination against overall influenza infection across all study participants. The original paper did not detect a significant protective effect of the vaccine from any source using either the unadjusted or adjusted model (Table 2.8). We estimated overall VE lower than both the unadjusted and adjusted estimates. Our method produced 95% CIs that were slightly wider for household VE, but narrower than those reported in the original study for community and overall VE.

Table 2.8: VE estimates based on hazard rate ratios and 95% confidence intervals from the HIVE Study

Model	$\mathbf{VE_H}$	$\mathbf{VE_C}$	$\mathbf{VE_O}$
Unadjusted (95% CI)	0.31 (-0.73, 0.73)	0.27 (-0.13, 0.54)	0.30 (-0.09, 0.55)
Adjusted (95% CI)	0.37 (-0.73, 0.77)	0.30 (-0.09, 0.55)	0.43 (-0.18, 0.72)

VE point estimates and 95% CIs from Ohmit *et al.* using both the unadjusted and adjusted hazard rate ratio. The adjusted models adjusted for age in months and documentation of high-risk health status [49]. VE_H, VE_C, and VE_O denote VE estimates against household, community, and overall transmission, respectively.

2.4 Discussion

We have presented a probability model and accompanying maximum likelihood procedure to estimate VE against household-acquired and community-acquired influenza infection from MH studies. Our method first estimates source-specific transmission parameters that characterize the daily probability of infection. We use these transmission parameters to estimate the probability of influenza infection throughout the study and estimate VE against transmission of influenza from the household and from the community. Previous methods

that estimate source-specific VE use final count data that do not take into account the time of infection [43, 91, 108]. Our approach improves upon these methods by incorporating time to event data, which allows for variation in influenza prevalence and timing of vaccination to be incorporated into estimates of VE. We used a stochastic agent-based simulation program to evaluate the bias and precision of our estimates.

Under our model assumptions, our method estimated the transmission parameters and VE close to the truth for two different vaccination scenarios (Tables 2.2 and 2.3). Transmission parameter estimates were very similar regardless of vaccination scenario. SEs were calculated empirically, using the Hessian matrix, and using a parametric bootstrap procedure. In all scenarios, bootstrap SEs were close to empirical SEs indicating that the bootstrap procedure performs well and is appropriate for the estimation of SE when analyzing real data (when estimation of SE empirically is not possible). Coverage probabilities of greater than or equal to 95% for all estimates of VE suggest that our method performs well. For estimates of VE against community-acquired influenza when all vaccinations occurred prior to the study, the coverage probability was slightly lower (92%) suggesting that our method produces CIs that are too narrow for this estimate.

We developed a similar likelihood method under the assumptions that the source of infection (household or community) is known. We found that the estimates and their standard deviations were similar to those we obtained without this assumption. Hence, we conclude that knowing the source of infection does not substantially improve the VE estimates.

Results from our sensitivity analyses suggest that our maximum likelihood approach provided estimates of source specific VE with small bias and SE when

the length of the latent and infectious periods are not constant and the mean is correctly specified. When the mean lengths of the latent and infectious periods were misspecified, our results suggest that caution should be used when using the maximum likelihood procedure to estimate community-acquired influenza when vaccination occurs during the study, as the estimates may be moderately biased. However, the misspecification of the mean lengths of the latent and infectious periods had little impact on SE of VE estimates.

Allowing the prevalence of influenza infection to differ in a randomly sampled cohort compared to the overall population had little impact on the bias of source-specific VE estimates (Table 2.5) indicating that the maximum likelihood approach is robust to differences in the prevalence of influenza between the study cohort and overall population. To investigate the bias of VE estimates when the sample population is a small fraction of the overall population, we performed an additional sensitivity analysis in which we simulated an overall population of 10,100 people with a sample population of 1,000 people (results not shown). We saw no change in the amount of bias when the cohort was a smaller fraction of the overall population. Under the assumption that the cohort is a random sample of the overall population, we would expect the results to be similar regardless of the size of the overall population relative to the cohort.

It is well known that not all individuals infected with influenza develop an ARI [74]. However, little is known about the proportion of asymptomatic individuals in a given influenza season. One study estimated that approximately 67% of individuals infected with influenza develop symptoms [74], while other studies have estimated that as few as 23% [114] or as many as 84% [115] of influenza infected individuals develop symptoms. Despite the lack of symptoms, asymptomatic individuals are still infectious; however, less

so than symptomatic individuals [74]. Little is known about the relative infectiousness of asymptomatic individuals compared to symptomatic individuals. While asymptomatic individuals are considered less infectious because they are not shedding as much virus as symptomatic individuals [116], asymptomatic individuals may make more contacts while infectious than their symptomatic counterparts because they do not realize they are infected. Due to the many unknowns surrounding asymptomatic influenza infections, we did not include asymptomatic individuals in our sensitivity analyses. It will be important in future work to assess the impact of asymptomatic individuals in the population on VE estimates.

We applied our method to data from the HIVE study during the 2012-2013 influenza season [49] (Table 2.7). Our VE point estimates for household and overall influenza infection were lower than the point estimates found in the original study (Table 2.8) using unadjusted and adjusted hazard rate ratios. This difference in point estimates is likely due to the fact that our method does not require source of infection to be known, and in the case of the adjusted estimates, does not control for potential confounders, such as health status and age. Our estimate of community VE was similar to the original estimates, but we were able to detect a significant protective effect of vaccination against community-acquired influenza infection across all study participants (Table 2.7). Ohmit *et al.* found no significant protective effect of vaccination for overall, household-acquired, and community-acquired influenza infection across all study participants using the hazard rate ratio (Table 2.8) [49].

Our probability model makes many simplifying assumptions about influenza disease progression. In future work, we plan to relax some of our model assumptions to more realistically model the influenza disease process. First, we

would like to allow for strata within the population, particularly age groups. Previous research indicates that transmission of influenza is different from child to child, child to adult, adult to child, and adult to adult [117]. Additionally, age has been identified as an important risk factor associated with influenza transmission in which children and the elderly are more susceptible to infection than young adults. Very young children and older adults are also more susceptible to complications from infection. The addition of strata requires additional assumptions about the contact patterns of individuals in the same stratum and between strata. Previous studies have found that contacts made by children and adolescents are more assortative than other age groups. The same study found that individuals aged 55 years and older had the least assortative contact patterns [118]. We plan to extend our method to incorporate these additional contact patterns. We also plan to use stratification to reduce confounding bias.

Second, we plan to extend our probability model for the all-or-none protection vaccination model in which a proportion of vaccinated individuals are completely protected from infection and the remaining vaccinated individuals are not protected at all [43, 44].

Finally, we plan to analyze data from different influenza seasons and different settings to better determine the effectiveness of influenza vaccination against household transmission compared to community transmission. One study during the 2010-2011 influenza season found that VE against household and community transmission were different [48]. However, more recent studies have not observed this difference in VE against household- and community-acquired influenza infection [49, 97]. Further research is required to elucidate the impact of contact dynamics within populations of households on influenza

VE; however, our model and maximum likelihood procedure provide a framework to begin distinguishing influenza VE from different sources of infection.

Chapter 3

A dynamic model for evaluation of bias of estimates of influenza vaccine effectiveness from observational studies

3.1 Introduction

3.1.1 Background

Influenza vaccination is required each year due to changes in the influenza viruses (antigenic drift) and waning immunity from vaccination. A new influenza vaccine is produced annually requiring influenza vaccine effectiveness (VE) to be estimated each year [107] using observational studies as RCTs are no longer considered ethical [11]. Observational studies usually produce biased VE estimates, thus we aim to compare VE estimates from four different study designs to (1) better characterize the conditions under which estimates of VE may be biased, (2) evaluate the magnitude and direction of the bias, and (3)

compare the bias of VE estimates from different study designs.

3.1.2 Study designs

We will evaluate the bias of influenza VE estimates produced by ASC, PSC, TN, and TCC studies. The study population in each of these studies consists of individuals who receive most of their medical care at a single clinic or network of clinics and consent to being inlcluded in the study. In an ASC study, a person is tested for influenza by study personnel when they develop and report symptoms of ARI regardless of whether or not they seek medical care for their symptoms. In PSC, TN, and TCC studies, when an individual develops an ARI, s/he may seek medical care at a clinic for treatment of the ARI and may be tested for influenza infection. Definitions of cases and non-cases/controls for each study design are shown in Table 3.1.

3.1.3 Sources of Bias

In this section we discuss the sources of bias that may be present in each type of study. For convenience, we use the acronyms FARI and NFARI for inFluenza and Non-inFluenza ARI, respectively.

A. **Vaccination affects the probability of NFARI:** Vaccination may modify the probability of developing NFARI. As a result, too many or too few vaccinated persons may be classified as controls/non-cases. Specifically, an assumption underlying the validity of TN studies is that vaccination does not effect of the probability of developing NFARI [45]. Recent studies to test this assumption have produced conflicting results [45, 66, 67].

Table 3.1: Definitions of cases and controls/non-cases for observational studies.

Study design	Cases	Controls/Non-cases
ASC	Individuals in the cohort who reported an ARI and test positive for influenza	All other members of the cohort
PSC	Individuals in the cohort who seek medical care for an ARI and test positive for influenza	All other members of the cohort
TN	Members of the study population who seek medical care for an ARI and test positive for influenza	Members of the study population who seek medical care for an ARI and test negative for influenza
TCC	Members of the study population who seek medical care for an ARI and test positive for influenza	Randomly selected individuals from the study population who did not develop an ARI throughout the study

ASC - active surveillance cohort, PSC - passive surveillance cohort, TN - test-negative, and TCC - traditional case-control.

B. **Confounding bias due to the presence of a covariate (e.g. health status) that is related to both the probability of being vaccinated and the probabilities of FARI and NFARI:** Health status may be associated with the probability of being vaccinated as frail persons may be more likely to be vaccinated because they are considered at higher risk for influenza infection. On the other hand, healthy persons may be more likely to be vaccinated to preserve their good health. Health status may also be associated with the probabilities of developing FARI and NFARI as frail persons may be more likely to have an ARI.

C. **Vaccination may modify the probability of seeking medical care:** A person who is vaccinated may have a different probability of seeking medical care for FARI compared to an unvaccinated person due to a reduction of

symptom severity from vaccination [119–121]. In PSC and TCC studies, cases are selected from individuals who seek medical care, while controls are not, which may confound the effect of vaccination [39]. A TN study attempts to control for this by selecting both cases and controls from persons seeking medical care for an ARI. This source of bias is removed from ASC studies because cases and non-cases can be identified regardless of whether a participant seeks medical care for an ARI.

D. **Confounding bias due to the presence of a covariate (e.g. health awareness) that is related to both the probabilities of vaccination and seeking medical care:** A person's health awareness may be associated with the probability of being vaccinated as a person with high health awareness may be more likely to be vaccinated. That same person may also be more likely to seek medical care if they develop an ARI.

E. **Misclassification bias:** Influenza diagnostic tests are not 100% sensitive and specific, resulting in false-positive or false-negative test results. Additionally, vaccination status may be misclassified. This source of bias may be present in all four study designs.

3.1.4 Outcomes of Interest

Previous work has shown that estimates of VE may change depending on the outcome of interest [99]. We will consider assessing VE with respect to two different outcomes of interest: SI and MAI. SI is defined as influenza infection resulting in an ARI. Although SI and FARI are identical concepts, SI is considered the true outcome (e.g., an outcome against which VE is estimated),

whereas FARI is an observed outcome. MAI is defined as an influenza infection resulting in an ARI for which a person seeks medical care.

TN, TCC, and PSC studies all require individuals with FARI to seek medical care to be considered a case thus, we expect these types of studies to provide estimates of VE against MAI. We will also assess the bias of estimates from these studies when the outcome of interest is SI. Evaluating VE against SI is important as influenza patients who do not seek medical care add to the burden of disease. Additionally, estimated VE against MAI may be misinterpreted by the media and public as VE against SI. Therefore, we are interested in assessing the validity of estimates of VE against SI from studies that are designed to provide estimates of VE against MAI. ASC studies usually do not collect information on seeking care, therefore we will only evaluate the bias of the resulting estimates of VE against SI.

3.1.5 Objectives of this Work

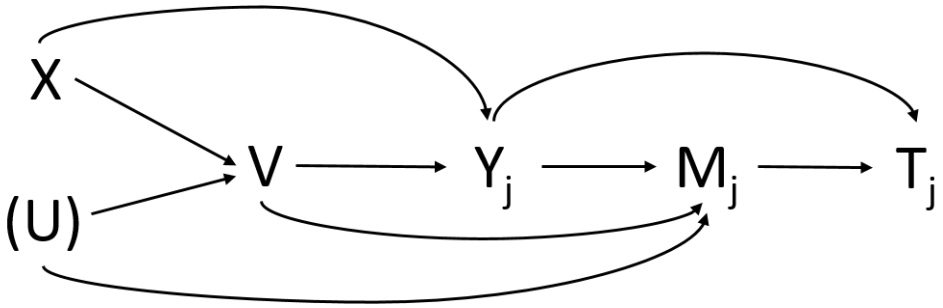
In this work, we present a dynamic model that allows the intensities of FARI and NFARI to change over time and allows for the possibility of developing more than one ARI in a season. The model incorporates factors that may affect the probabilities of vaccination, developing FARI and/or NFARI, and seeking medical care for ARI. We use the model to evaluate and compare the bias of VE estimates from four different observational study designs against two outcomes of interest. This work extends a previous paper [99] which used a static model for evaluating the bias of VE estimates from the two types of case-control studies.

3.2 Methods

3.2.1 Model Description

We present a dynamic model consisting of five steps. Below we define the model steps, the associated variables (Table 3.2), and the probabilities determining each variable's distribution (which may depend on variables from previous steps). Table 3.3 lists all of the model parameters. All variables are defined for each member of the study population. Within this model we allow some variables to change over time. We consider each time unit to be one week. In Figure 3.1 we present a directed acyclic graph [122, 123] to illustrate the possible sources of confounding and bias present in studies designed to evaluate influenza VE.

Figure 3.1: Causal graph of influenza vaccine studies with covariates.



X = health status, (U) = health awareness (unobserved), V = vaccination status, Y_j = ARI status in week j, M_j = seeking medical care for ARI in week j, and T_j = influenza test result in week j, where $j = 1, \ldots, J$ and J is the total number of weeks in the study.

Step 1: Covariates. We assume that people within the population can be classified with a health status of either "healthy" or "frail" and a health awareness of either "high" or "low". We define an observable binary variable X, where

$X = 1$ for a "healthy" person and an unobserved binary variable U, where $U = 1$ for a person with high health awareness. Let $\pi_{xu} = \mathbb{P}(X = x, U = u)$.

Table 3.2: Variables in the model.

Variable	Definition	Values
X	Health Status	0 - frail person 1 - healthy person
U	Health Awareness (unobserved)	0 - low health awareness 1 - high health awareness
V	Vaccination Status	0 - unvaccinated 1 - vaccinated
Y_j	Influenza/non-influenza ARI status at week j	0 - no ARI 1 - non-influenza ARI 2 - influenza ARI
M_j	Seeking medical care for ARI at week j	0 - no 1 - yes
T_j	Result of test for influenza infection at week j	0 - negative 1 - positive

Step 2: Vaccination. We consider the scenario where everyone is either effectively vaccinated prior to the study or remains unvaccinated throughout the study. A person is considered effectively vaccinated if they received the vaccine at least 14 days prior to the study onset. We define a binary variable V, where $V = 1$ if a person is effectively vaccinated. The probability of vaccination may depend on X and U. Let $\alpha_{xu} = \mathbb{P}(V = 1|X = x, U = u)$.

For the remainder of the steps, it will be convenient to define a *standard person* as a person who is unvaccinated, healthy, and has high health awareness (i.e., a person with $V = 0$, $X = 1$, and $U = 1$).

Step 3: Influenza and non-influenza ARI. During the influenza season, a person may become infected with an influenza virus and develop FARI. Regardless of influenza infection, a person may develop one or more NFARIs. We

assume that an individual can only become infected with influenza once during the study period and a person can have no more than one NFARI per week, but there is no limit on the total number of NFARIs. We define a variable Y_j for the illness/infection status in week j as follows: $Y_j = 0$ for no ARI, $Y_j = 1$ for NFARI, and $Y_j = 2$ for FARI. If a person has both NFARI and FARI in the same week, we consider them as FARI (i.e., $Y_j = 2$). The distribution of Y_j may depend on the person's vaccination (V) and health (X) status. We denote the probability of NFARI in week j as $\beta_{jvx} = \mathbb{P}(Y_j = 1|V = v, X = x), j = 1, \ldots, J$. We denote the probability of NFARI for a *standard person* β_{j01} and specify the value of β_{j01} for all j. For all other persons we define multipliers for β, θ_β and ϕ_β, where θ_β is the multiplier when $V = 1$ and ϕ_β is the multiplier when $X = 0$. Then, the probabilities of NFARI in week j for non-standard persons are $\beta_{j11} = \beta_{j01} \cdot \theta_\beta$, $\beta_{j00} = \beta_{j01} \cdot \phi_\beta$, and $\beta_{j10} = \beta_{j01} \cdot \theta_\beta \cdot \phi_\beta$.

We denote the probability of FARI in week j as $\gamma_{jvx} = \mathbb{P}(Y_j = 2|V = v, X = x), j = 1, \ldots, J$. Similarly to the probability of NFARI, the probability of FARI for standard persons is denoted by γ_{j01} and we specify the value of γ_{j01} for all j. We define multipliers for γ, θ_γ and ϕ_γ, where θ_γ is the multiplier when $V = 1$ and ϕ_γ is the multiplier when $X = 0$. The probabilities of FARI in week j for non-standard persons are $\gamma_{j11} = \gamma_{j01} \cdot \theta_\gamma$, $\gamma_{j00} = \gamma_{j01} \cdot \phi_\gamma$, and $\gamma_{j10} = \gamma_{j01} \cdot \theta_\gamma \cdot \phi_\gamma$. We also assume $\beta_{jvx} + \gamma_{jvx} \leq 1$ in week j for all v and x.

Step 4: Seeking medical care for ARI. A person with an ARI may seek medical care. We define a binary variable M_j for whether or not a person seeks medical care for ARI in week j. The probability of seeking medical care depends on Y_j, as only those individuals who have an ARI may seek medical care, and it may be different for FARI and NFARI patients. This probability may also depend on V and U. We assume that the conditional probability of M given Y is

fixed over time and $\mathbb{P}(M_j = 1|Y_j = 0) = 0$.

The probability of a person seeking medical care in week j for NFARI is denoted as $\delta_{1vu} = \mathbb{P}(M_j = 1|Y_j = 1, V = v, U = u), v = 0, 1; u = 0, 1; j = 1, \ldots, J$. For the *standard person*, the probability of seeking care for NFARI in week j, $\delta_{101} = \mathbb{P}(M_j = 1|Y_j = 1, V = 0, U = 1)$, is an input parameter. We define multipliers for δ_1, θ_{δ_1} and μ_{δ_1}, where θ_{δ_1} is the multiplier when $V = 1$ and μ_{δ_1} is the multiplier when $U = 0$, such that $\delta_{111} = \delta_{101} \cdot \theta_{\delta_1}$, $\delta_{100} = \delta_{101} \cdot \mu_{\delta_1}$, and $\delta_{110} = \delta_{101} \cdot \theta_{\delta_1} \cdot \mu_{\delta_1}$.

Similar notation is used for FARI. The probability of seeking care for FARI in week j is $\delta_{2vu} = \mathbb{P}(M_j = 1|Y_j = 2, V = v, U = u), v = 0, 1; u = 0, 1; j = 1, \ldots, J$. For the *standard person*, $\delta_{201} = \mathbb{P}(M_j = 1|Y_j = 2, V = 0, U = 1)$ is an input parameter. The multipliers for δ_2 are θ_{δ_2}, when $V = 1$, and μ_{δ_2}, when $U = 0$, such that δ_{2vu} can be expressed in terms of δ_{201} and the multipliers.

Step 5: Testing for influenza infection. We assume that each person who seeks medical care for ARI is tested for influenza infection. We define a binary variable T_j as the test result for a person with ARI in week j, where $T_j = 1$ for an influenza positive test result. We assume the probability of testing positive for influenza given a person's influenza infection status does not depend on any other factors. We denote $\tau_y = \mathbb{P}(T_j = 1|Y_j = y)$, $y = 1, 2$. Note that τ_1 is one minus the test's specificity and τ_2 is the test's sensitivity. Since only persons with ARIs are tested, τ_0 is irrelevant.

Assumptions

We make several assumptions in the model: *(a)* Each person is assigned two binary covariates: health status and health awareness, which are constant over time, *(b)* everyone is either effectively vaccinated prior to the study or remains

Table 3.3: Parameters in the model.

Parameters	Definition	Input Value	Details
π_{xu}	$\mathbb{P}(X=x, U=u)$	$\pi_{11}=0.40$ $\pi_{10}=0.40$ $\pi_{01}=0.10$ $\pi_{00}=0.10$	
α_{xu}	$\mathbb{P}(V=1\|X=x, U=u)$	$\alpha_{11}=0.60$ $\alpha_{10}=0.30$ $\alpha_{01}=0.90$ $\alpha_{00}=0.45$	
β_{jvx}	$\mathbb{P}(Y_j=1\|V=v, X=x)$	see Table B1	$\beta_{j11}=\beta_{j01}\cdot\theta_\beta$
θ_β	multiplier for β when $V=1$	see Table 3.7	$\beta_{j00}=\beta_{j01}\cdot\phi_\beta$
ϕ_β	multiplier for β when $X=0$	see Table 3.7	$\beta_{j10}=\beta_{j01}\cdot\theta_\beta\cdot\phi_\beta$
γ_{jvx}	$\mathbb{P}(Y_j=2\|V=v, X=x)$	see Table B1	$\gamma_{j11}=\gamma_{j01}\cdot\theta_\gamma$
θ_γ	multiplier for γ when $V=1$	see Table 3.7	$\gamma_{j00}=\gamma_{j01}\cdot\phi_\gamma$
ϕ_γ	multiplier for γ when $X=0$	see Table 3.7	$\gamma_{j10}=\gamma_{j01}\cdot\theta_\gamma\cdot\phi_\gamma$
δ_{1vu}	$\mathbb{P}(M_j=1\|Y_j=1, V=v, U=u)$	$\delta_{101}=0.25$	$\delta_{111}=\delta_{101}\cdot\theta_{\delta_1}$
θ_{δ_1}	multiplier for δ_1 when $V=1$	see Table 3.7	$\delta_{100}=\delta_{101}\cdot\mu_{\delta_1}$
μ_{δ_1}	multiplier for δ_1 when $U=0$	see Table 3.7	$\delta_{110}=\delta_{101}\cdot\theta_{\delta_1}\cdot\mu_{\delta_1}$
δ_{2vu}	$\mathbb{P}(M_j=1\|Y_j=2, V=v, U=u)$	$\delta_{201}=0.40$	$\delta_{211}=\delta_{201}\cdot\theta_{\delta_2}$
θ_{δ_2}	multiplier for δ_2 when $V=1$	see Table 3.7	$\delta_{200}=\delta_{201}\cdot\mu_{\delta_2}$
μ_{δ_2}	multiplier for δ_2 when $U=0$	see Table 3.7	$\delta_{210}=\delta_{201}\cdot\theta_{\delta_2}\cdot\mu_{\delta_2}$
τ_y	$\mathbb{P}(T=1\|Y=y)$	$\tau_1=0$ $\tau_2=1$	

For a *standard person* ($V=0, X=1, U=1$), β_{j01} and γ_{j01} represent the probabilities of contracting a NFARI or FARI in week j, respectively, $j=1,\ldots,J$, and δ_{101} and δ_{201} represent the probability of seeking care for NFARI and FARI, respectively. These probabilities, as well as all π_{xu}, $x=0,1; u=0,1$, all α_{xu}, $x=0,1; u=0,1$, τ_1, τ_2, and all multipliers ($\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$), are input parameters.

unvaccinated throughout the study, and vaccination status is determined without error, *(c)* everyone in the study population is susceptible at the beginning of the study, *(d)* a person can only have one FARI during the season, *(e)* a person can have at most one NFARI per week, *(f)* the probabilities of FARI and NFARI do not depend on a person's health awareness given his/her health status, *(g)* the probability of seeking medical care does not depend on a person's health

status given his/her health awareness, *(h)* influenza test specificity and sensitivity do not depend on health status, health awareness, or vaccination status, given a person's influenza infection status, *(i)* every person who seeks medical care is tested for influenza infection, and *(j)* a person with no ARI does not seek care (i.e., if $Y_j = 0$, then $M_j = 0$).

3.2.2 True VE

The bias of a VE estimate is the difference between the expected value of the estimate and the true VE. True VE is defined as one minus RR, where RR is the probability of the outcome given vaccination divided by the probability of the same outcome given no vaccination when vaccination is random (i.e., the probability of vaccination does not depend on any covariates). Previously, we showed that the true VEs against SI and MAI may differ [99]. Therefore, we evaluate the true VE for each of the two outcomes of interest. Explicit expressions and derivations of true VE can be found in Appendices B.1.1 and B.1.2.

True VE Against SI

A person is considered a true case of SI if s/he develops an influenza ARI during the study. True VE against SI (VE_{TSI}) is

$$VE_{TSI} = 1 - RR_{TSI} \quad \text{where} \quad RR_{TSI} = \frac{\mathbb{P}(\text{contracting SI}|V=1)}{\mathbb{P}(\text{contracting SI}|V=0)},$$

where $\mathbb{P}(\text{contracting SI}|V=v) = \sum_{j=1}^{J} \mathbb{P}(Y_j = 2|V=v)$, $v = 0, 1$;
$j = 1, \ldots, J$.

True VE Against MAI

A person is considered a true case of MAI if s/he develops an influenza ARI during the study and seeks medical care for this ARI. True VE against MAI (VE_{TMAI}) is

$$VE_{TMAI} = 1 - RR_{TMAI} \quad \text{where} \quad RR_{TMAI} = \frac{\mathbb{P}(\text{classified as MAI}|V=1)}{\mathbb{P}(\text{classified as MAI}|V=0)},$$

where $\mathbb{P}(\text{classified as MAI}|V=v) = \sum_{j=1}^{J} \mathbb{P}(Y_j = 2, M_j = 1|V=v)$, $v = 0, 1; j = 1, \ldots, J$.

3.2.3 VE Estimates

In ASC and PSC studies, VE is estimated as $\hat{VE} = 1 - \hat{RR}$, where $\hat{RR}$ is based on sample proportions. In TN and TCC studies, VE is estimated as $1 - OR$, where OR is the odds ratio comparing the odds of vaccination in cases and controls [46, 55–57]. In this work we consider only unadjusted estimates of VE because we are interested in characterizing bias rather than in methods to adjust for it. Current VE estimates are not specifically designed to estimate VE against SI or MAI. Therefore, we will not account for the outcome of interest when we develop expressions for these estimates. The outcome of interest will be accounted for when we evaluate the bias of these estimates (Section 3.3). Explicit expressions and derivations of estimates of VE from each study design can be found in Appendices B.2–B.4. We write the VE estimates in terms of probabilities; in actual studies, these probabilities are replaced by the corresponding proportions. Therefore, the 'VE estimates' we calculate are the expected values of the actual estimates under our assumptions.

For each person, $\mathbf{Y_J} = (Y_1, \ldots, Y_J)$ and $\mathbf{M_J} = (M_1, \ldots, M_J)$ are the arrays of Y_j and M_j values, respectively, from week 1 to week J.

ASC Studies

For a person to be considered a case, they must have an ARI ($Y_j > 0$) and test positive ($T_j = 1$) in at least one week j. Since we do not assume perfect influenza test sensitivity and specificity, a person may test positive in more than one week (even though s/he can have true influenza only once).

The probability of being a case in an ASC study for a given v can be written as

$$\mathbb{P}(\text{Case}(ASC)|V = v) \quad = \quad \mathbb{P}(\bigcup_{j=1}^{J} [\{Y_j = 1, T_j = 1\} \cup \{Y_j = 2, T_j = 1\}])$$

Thus, the VE estimate from an ASC study is

$$\hat{VE}_{ASC} = 1 - \frac{\hat{\mathbb{P}}(\text{Case}(ASC)|V = 1)}{\hat{\mathbb{P}}(\text{Case}(ASC)|V = 0)}$$

PSC Studies

In a PSC study, a person is considered a case in week j if:

- they have an ARI in week j (i.e., $Y_j > 0$)
- they seek medical care for their ARI in week j (i.e., $M_j = 1$)
- they test positive for influenza infection (i.e., $T_j = 1$)

The probability of being a PSC case for a given v is the probability of being a case in at least one week.

$$\mathbb{P}(\text{Case}(PSC)|V=v) \;=\; \mathbb{P}(\bigcup_{j=1}^{J}[\{Y_j=1,M_j=1,T_j=1\}\cup\{Y_j=2,M_j=1,T_j=1\}])$$

Thus, the VE estimate from an PSC study is

$$\hat{VE}_{PSC}=1-\frac{\hat{\mathbb{P}}(\text{Case}(PSC)|V=1)}{\hat{\mathbb{P}}(\text{Case}(PSC)|V=0)}$$

TN Studies

We assume that a person is classified as a TN case or a TN control at her/his first ARI-related visit. This classification does not change, regardless of possible conflicting test results in future visits. A person is considered a case in week j, if:

- they did not seek medical care for any ARI prior to week j, so $M_k=0$ for every week $k=1,\ldots,j-1$ (i.e., $\mathbf{M_{j-1}=0}$)
- they seek medical care for their ARI in week j (i.e., $M_j=1$)
- they are diagnosed with FARI in week j (i.e., $T_j=1$)

The probability of being considered a case for a given v is:

$$\mathbb{P}(\text{Case}(TN,TCC)|V=v) \;=\; \sum_{j=1}^{J}\mathbb{P}(\mathbf{M_{j-1}=0},M_j=1,T_j=1|V=v)$$

A TN control in week j is defined the same as a TN case, except they test negative for influenza infection (i.e., $T_j = 0$). The probability of being considered a TN control for a given vaccination status v is

$$\mathbb{P}(\text{Control}(TN)|V = v) \quad = \quad \sum_{j=1}^{J} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, T_j = 0|V = v)$$

The VE estimate from a TN study is

$$\hat{VE}_{TN} = 1 - \hat{OR}_{TN},$$

where $\hat{OR}_{TN}$ is the odds ratio from Table 3.4.

Table 3.4: Final 2x2 table for a TN study.

	Vaccinated	**Unvaccinated**
Case	$\hat{\mathbb{P}}(\text{Case}(TN, TCC)|V = 1)$	$\hat{\mathbb{P}}(\text{Case}(TN, TCC)|V = 0)$
Control	$\hat{\mathbb{P}}(\text{Control}(TN)|V = 1)$	$\hat{\mathbb{P}}(\text{Control}(TN)|V = 0)$

TCC Studies

A TCC case is classified in the same way as a TN case . TCC controls are selected after the end of the study. A person is considered a TCC control if s/he did not have an ARI during the entire study period (i.e., $\mathbf{Y_J} = \mathbf{0}$).

The probability of being considered a TCC control for a given vaccination status v is

$$\mathbb{P}(\text{Control}(TCC)|V = v) \quad = \quad \mathbb{P}(\mathbf{Y_J} = \mathbf{0}|V = v)$$

The VE estimate from a TCC study is

$$\hat{VE}_{TCC} = 1 - \hat{OR}_{TCC},$$

where $\hat{OR}_{TCC}$ is the odds ratio from Table 3.5.

Table 3.5: Final 2x2 table for a TCC study.

	Vaccinated	**Unvaccinated**
Case	$\hat{\mathbb{P}}(\text{Case}(TN, TCC)\|V = 1)$	$\hat{\mathbb{P}}(\text{Case}(TN, TCC)\|V = 0)$
Control	$\hat{\mathbb{P}}(\text{Control}(TCC)\|V = 1)$	$\hat{\mathbb{P}}(\text{Control}(TCC)\|V = 0)$

3.2.4 Sources of Bias

The various sources of bias that may occur under our model and assumptions are listed in Table 3.6. We use labels to refer to these sources of bias throughout the Results and Discussion sections. Each source of bias can be attributed to deviation of a specific probability ratio from 1. Table 3.7 lists the probability ratio corresponding to each source of bias, the corresponding model parameter, and the range of values used for that parameter in our calculations and simulations.

Table 3.6: Sources of bias.

Label	**Source of Bias**
A	Vaccination affects the probability of NFARI
B1	Healthy persons have a lower probability of NFARI
B2	Healthy persons have a lower probability of FARI
BS	Healthy persons have a lower probability of FARI and NFARI. Health status has the same effect on the probabilities of both types of ARI.
C	Vaccination lowers the probability of seeking medical care in FARI patients (because of reduced symptom severity).
D	ARI patients with high health awareness have a higher probability of seeking medical care.
E	Misclassification of influenza infection status

In our calculations and simulations, we assumed perfect sensitivity and specificity. In other words, we will not account for bias E.

Table 3.7: Probability ratios corresponding to sources of bias.

Source of Bias	Probability Ratio	Definition	Parameter	Range
A	PR_A	$\mathbb{P}$(NFARI|Vacc)/$\mathbb{P}$(NFARI|Unv)	θ_β	0.5 - 2.0
B1	PR_{B1}	$\mathbb{P}$(NFARI|Frail)/$\mathbb{P}$(NFARI|Healthy)	ϕ_β	1.0 - 2.0
B2	PR_{B2}	$\mathbb{P}$(FARI|Frail)/$\mathbb{P}$(FARI|Healthy)	ϕ_γ	1.0 - 2.0
BS	PR_{BS}	Common value PR_{B1} and PR_{B2}	$\phi_\beta = \phi_\gamma$	1.0 - 2.0
C	PR_C	$\mathbb{P}$(SMC|FARI, Vacc)/$\mathbb{P}$(SMC|FARI, Unv)	θ_{δ_2}	0.5 - 1.0
D	PR_D	$\mathbb{P}$(SMC|Low HA)/$\mathbb{P}$(SMC|High HA)	$\mu_{\delta_1} = \mu_{\delta_2}$	0.5 - 1.0

PR - Probability ratio, Vacc - Vaccinated, Unv - Unvaccinated, FARI - Influenza ARI, NFARI - Non-influenza ARI, HA - Health awareness, SMC - Seeking medical care

3.2.5 Calculations

To evaluate the bias of VE estimates under different sources of bias (Table 3.6), we derived expressions of true and estimated VE from our model (appendix B.2). Using these expressions, we calculated the bias of VE estimates under various combinations of sources of bias by varying the values of the corresponding parameters (probability ratios, Table 3.7). When a source of bias was absent, we kept the corresponding probability ratio fixed at 1.0. Bias was defined as estimated VE minus true VE. For bias A, we considered that vaccination might increase or decrease the probability of NFARI, so θ_β varied from 0.5 to 2.0. For biases B1, B2, and BS, we allowed ϕ_β and ϕ_γ to vary between 0.5 and 1.0, since we expect healthy persons to have lower probabilities of ARI compared to frail persons. For bias C, θ_{δ_2} varied between 0.5 to 1.0, since we expect vaccination to reduce the probability of seeking medical care for FARI compared to NFARI. For bias D, we assume $\mu_{\delta_1} = \mu_{\delta_2}$, and their common value varied between 0.5

to 1.0 because we expect persons with high health awareness to have a higher probability of seeking medical care for both FARI and NFARI compared to persons with low health awareness. For each study design and each combination of sources of bias we determined the 5^{th}, 50^{th}, and 95^{th} quantiles of bias from 1,000 Monte Carlo simulations. For each simulation, values for the relevant probability ratio(s) were drawn from independent triangular distributions over the ranges specified in Table 3.7. The mode of each distribution was assumed to be 1. The true and estimated VE were calculated for each simulation. Values of input parameters used in the calculations can be found in Appendix B.5.

3.2.6 Simulations

To validate our calculated results, we used a stochastic simulation program to simulate data from each type of study design. Each simulation consisted of a population of 30,000 people. Each person was assigned a health status, health awareness, and vaccination status. A random sample of 5,000 and 10,000 people were assigned to be the active surveillance and passive surveillance cohorts, respectively. The remaining people in the population were eligible to be included in one of the case-control studies. The same parameter values used in the calculations were used in the simulation (see Appendix B.5). The parameter value corresponding to the source of bias being assessed was varied by 0.1 while keeping all other parameter values fixed. 1000 simulations were performed for each combination of parameter values. True VE and estimated VE were determined for each simulation. True VE was simulated under random vaccination. VE estimates were obtained from the observed frequencies of cases and controls by vaccination status. The mean true and estimated VE

values over all simulations were compared to the calculated values. All simulations were performed in R 3.3.1 [112].

3.2.7 Sensitivity Analyses

To assess whether the value of true VE had an effect on the magnitude of bias, we performed our calculations using different values of true VE (0.2, 0.4, 0.6, and 0.8). We also varied the probabilities related to health status, health awareness, and vaccination to determine if and how the choice of these probabilities impacts the bias of VE estimates. Finally, we varied the value of influenza test sensitivity and specificity. For each sensitivity analysis, bias was calculated in the presence of each source of bias for every study design.

Alternative Values of True VE

To evaluate the magnitude of bias for different values of true VE, we performed our calculations when the true VE was 0.2, 0.4, 0.6, and 0.8.

Alternative Probabilities of (X, U)

We selected two additional sets of input probabilities of X and U to assess the dependency of the amount of bias of VE estimates on our chosen probabilities of health status and health awareness. Originally, we assumed independence between the probabilities of X and U (i.e., a person's health status did not affect their health awareness). In the first set of alternative probabilities, we selected values such that the probability of having high health awareness ($U = 1$) given a healthy health status ($X = 1$) is greater than the probability of having high health awareness given a frail health status ($X = 0$). In the second set of alternative probabilities, we selected probabilities for the opposite scenario (i.e., the

pobability of having high health awareness given a healthy health status is *less* than the probability of having high health awareness given a frail health status). Table 3.8 shows the input probabilities for the original scenario (independence) and the two alternative scenarios.

Table 3.8: Alternative X and U probabilities.

Scenario	Parameters	Input Values
Independent	π_{11}	0.40
	π_{10}	0.40
	π_{01}	0.10
	π_{00}	0.10
$\mathbb{P}(U=1\|X=1) > \mathbb{P}(U=1\|X=0)$	π_{11}	0.44
	π_{10}	0.36
	π_{01}	0.06
	π_{00}	0.14
$\mathbb{P}(U=1\|X=1) < \mathbb{P}(U=1\|X=0)$	π_{11}	0.32
	π_{10}	0.48
	π_{01}	0.18
	π_{00}	0.02

Alternative Probabilities of Vaccination

We selected two additional sets of input probabilities of vaccination to assess the dependency of the amount of bias of VE estimates on our chosen probabilities of vaccination. Originally, we selected values such that the probability of being vaccinated was multiplicatively higher in individuals with frail health status ($X = 0$), high health awareness ($U = 1$), or both ($X = 0, U = 1$). In the first set of alternative probabilities we selected values such that the probability of being vaccinated was lower in people with frail health status ($X = 0$), low health awareness ($U = 0$), or both ($X = 0, U = 0$). For the second set of alternative probabilities, we selected values such that persons with frail health

status and high health awareness had a much higher probability of vaccination as we expect this group to have the highest probability of vaccination. Table 3.9 shows the alternative input probabilities.

Table 3.9: Alternative probabilities of vaccination.

Scenario	Parameters	Input Values
Original	α_{11}	0.60
	α_{10}	0.30
	α_{01}	0.90
	α_{00}	0.45
Alternative 1	α_{11}	0.60
	α_{10}	0.30
	α_{01}	0.30
	α_{00}	0.15
Alternative 2	α_{11}	0.60
	α_{10}	0.20
	α_{01}	0.95
	α_{00}	0.05

Sensitivity and Specificity

We varied the values of influenza test sensitivity and specificity to assess how deviations from our assumption of 100% sensitivity and specificity impacted influenza VE estimates. We assessed four combinations of sensitivity and specificity: (0.9, 0.9), (0.9, 1.0), (1.0, 0.9), and (1.0, 1.0). For each combination of sensitivity and specificity, the median absolute value of bias was calculated in the presence of sources of bias A, BS, C, and D (separately) for every study design.

3.3 Results

We evaluated bias of VE estimates from cohort and case-control studies in the presence of the sources of bias listed in Table 3.6. Table 3.10 shows the 5^{th}, 50^{th}, and 95^{th} quantiles of bias for each study design first, for each source of bias separately and then, for combinations of source of bias. We define several terms to aid in our evaluation of the magnitude of bias: little/small or no bias indicates an absolute bias of less than 0.05, moderate bias indicates absolute bias greater than or equal to 0.05 and less than 0.10, substantial bias indicates absolute bias greater than or equal to 0.10 and less than 0.20, and severe bias indicates absolute bias of 0.20 or more.

When no sources of bias are present, all study designs, except the TCC (bias=0.02), produce unbiased estimates of VE (Table 3.10). VE estimates from both the TN and TCC studies are based on the odds ratio rather than on the risk ratio, however, the TN-based estimate is unbiased even if the rare disease assumption does not hold [99]. Under bias A, both cohort studies produced unbiased VE estimates, while the case-control studies produce VE estimates with a wide range of bias (TN 90% Interval: (-0.34, 0.22), TCC 90% Interval: (-0.15, 0.10)). Interestingly, the direction of bias of VE estimates from the TN and TCC estimates is opposite as the probability ratio varies (Figure 3.2). Under bias B1, VE estimates from both cohort studies are unbiased. Estimates from the TN and TCC may suffer from small positive bias (TN: (0.00, 0.05), TCC: (0.00, 0.02)). Under bias B2, VE estimates from all four studies may suffer from small bias (ASC: (-0.05, 0.00), PSC: (-0.05, 0.00), TN: (-0.05, 0.00), TCC: (-0.03, 0.02)). Under bias BS, VE estimates from TN studies are unbiased, while estimates from the other three study designs may still suffer from small bias. Under

bias C, estimates of VE against SI are unbiased from ASC studies, but may be severely biased from the other three studies (PSC: (0.01, 0.23), TN: (0.01, 0.23), TCC: (0.03, 0.24)). Estimates of VE against MAI are unbiased from PSC and TN studies, while estimates from TCC studies have little bias. As mentioned earlier, we do not evaluate the bias of VE estimates against MAI from ASC studies (represented as a blank plot in Figure 3.2). Under bias D, estimates of VE from ASC and TN studies are unbiased. Estimates of VE from PSC and TCC studies may suffer from substantial bias (PSC: (-0.10, 0.00), TCC: (-0.07, 0.02)). All of our calculated results were validated using our stochastic simulation program (results not shown).

Next, we evaluated the bias of VE estimates when multiple sources of bias were present simultaneously (Table 3.10). We expect biases B and D to occur most often because they represent confounding due to an association between the covariates and the likelihood of illness and seeking medical care. Furthermore, we add the assumption that health status has the same effect on the probabilities of FARI and NFARI; therefore, we consider the bias BS rather than biases B1 and B2. The presence of biases A and C is more controversial. Therefore, biases BS and D were included in all scenarios, while the presence of biases A and C was varied. For these reasons, we looked at four scenarios: BS and D only; BS, D, and A; BS, D, and C; BS, D, A, and C.

- In the presence of BS and D only, ASC and TN studies produced estimates with little or no bias, while PSC and TCC studies produced estimates with moderate to substantial bias.

- With the addition of bias A (BS, D, A), ASC studies still produced estimates with little or no bias, while the remaining study designs produced

estimates with bias that ranged from little to severe (PSC 90% Interval: (-0.14, -0.01), TN: (-0.35, 0.22), TCC: (-0.26,0.06)).

- In the presence of BS, D, and C, the ASC design produced estimates of VE against SI with little or no bias, while the remaining study designs produced estimates of VE against SI with bias that ranged from little to severe (PSC: (-0.07, 0.20), TN: (0.01, 0.24), TCC: (-0.06,0.21)). The TN design produced unbiased estimates of VE against MAI. PSC and TCC studies produced estimates of VE against MAI that ranged from unbiased to substantially biased (PSC: (-0.11, -0.01), TCC: (-0.10,0.00)).

- When all four sources of bias were present (BS, D, A, C), the ASC design produced estimates of VE against SI with little or no bias, estimates of VE against SI and MAI from the other three study designs suffered from moderate to severe bias.

Table 3.10: Bias of estimates of VE against SI and MAI from observational studies

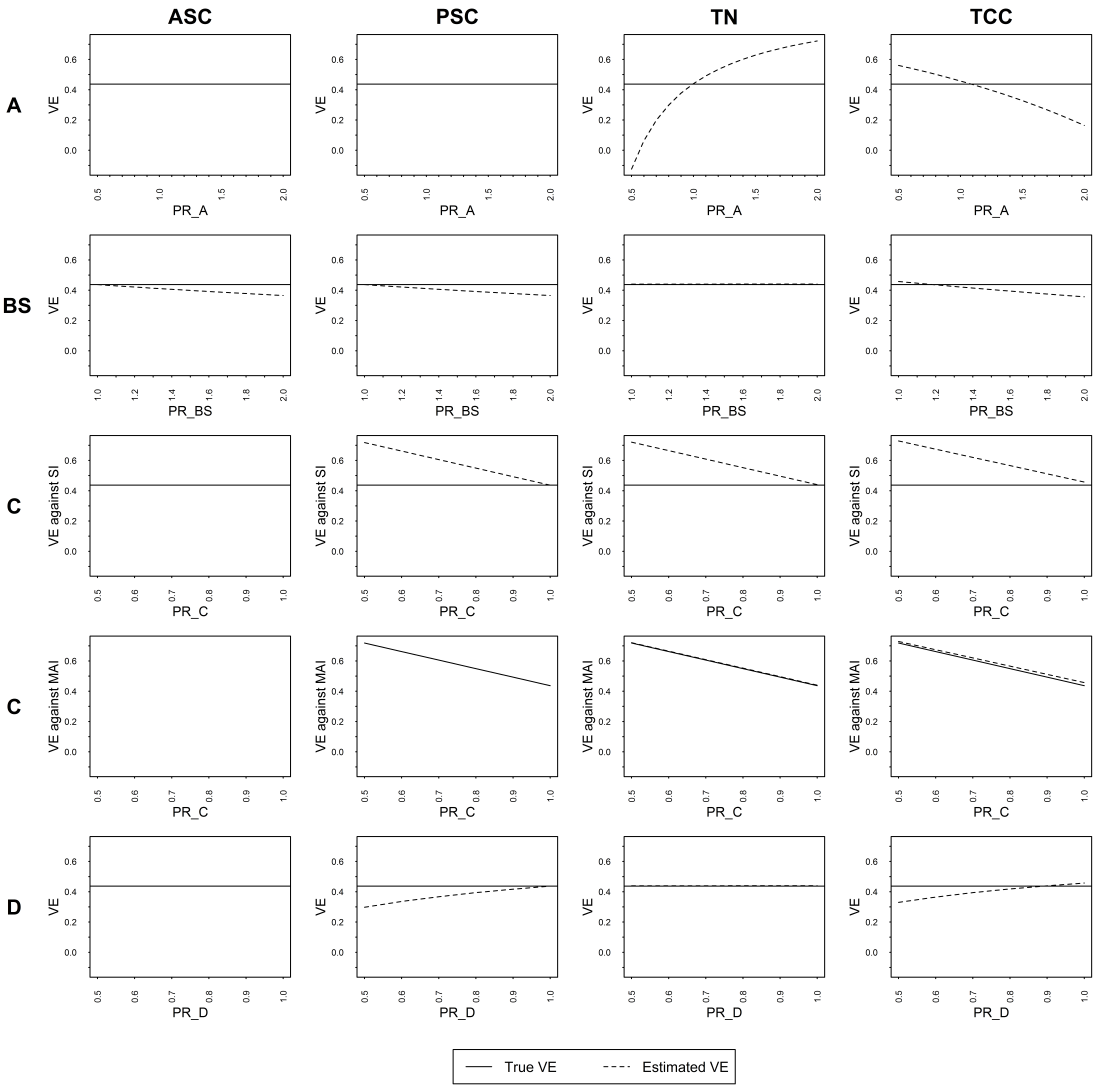
Source of Bias[a]	Outcome of Interest	ASC			PSC			TN			TCC		
		Q(5)[b]	Q(50)	Q(95)	Q(5)	Q(50)	Q(95)	Q(5)	Q(50)	Q(95)	Q(5)	Q(50)	Q(95)
None	SI & MAI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.02
A	SI & MAI	0.00	0.00	0.00	0.00	0.00	0.00	-0.34	0.01	0.22	-0.15	0.02	0.10
B1	SI & MAI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	-0.03	0.00	0.02
B2	SI & MAI	-0.05	-0.02	0.00	-0.05	-0.02	0.00	-0.05	-0.01	0.00	-0.03	0.00	0.02
BS	SI & MAI	-0.06	-0.02	0.00	-0.05	-0.02	0.00	0.00	0.00	0.00	-0.05	0.00	0.02
C	SI	0.00	0.00	0.00	0.01	0.10	0.23	0.01	0.10	0.23	0.03	0.12	0.24
C	MAI	–	–	–	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.02
D	SI & MAI	0.00	0.00	0.00	-0.10	-0.04	0.00	0.00	0.00	0.00	-0.07	-0.02	0.02
BS,D	SI & MAI	-0.05	-0.02	0.00	-0.14	-0.06	-0.01	0.00	0.00	0.00	-0.12	-0.05	0.00
BS,D,A	SI & MAI	-0.05	-0.02	0.00	-0.14	-0.07	-0.01	-0.35	0.01	0.22	-0.26	-0.05	0.06
BS,D,C	SI	-0.05	-0.02	0.00	-0.07	0.05	0.20	0.01	0.11	0.24	-0.06	0.06	0.21
BS,D,C	MAI[c]	–	–	–	-0.11	-0.05	-0.01	0.00	0.00	0.00	-0.10	-0.04	0.00
BS,D,A,C	SI	-0.05	-0.02	0.00	-0.08	0.05	0.20	-0.19	0.12	0.30	-0.14	0.06	0.23
BS,D,A,C	MAI	–	–	–	-0.12	-0.05	-0.01	-0.27	0.00	0.18	-0.21	-0.04	0.05

[a] Sources of bias: A - vaccination affects the probability of NFARI, B1 - Healthy persons have a lower probability of NFARI, B2 - Healthy persons have a lower probability of FARI, BS - Healthy persons have a lower probability of NFARI and FARI, C - Vaccination lowers the probability of seeking medical care in FARI patients (because of reduced symptom severity), and D - ARI patients with high health awareness have a higher probability of seeking medical care. Bias = estimated VE - true VE. The sign of bias indicates the direction of the difference between the estimated and true VE. A negative sign corresponds to underestimation while a positive bias indicates overestimation.

[b] Q(5)=5th quantile, Q(50)=50th quantile (median), and Q(95)=95th quantile.

[c] We do not estimate VE against MAI for ASC studies.

Figure 3.2: Plots of VE estimates from ASC, PSC, TN, and TCC studies compared to true VE for each source of bias alone



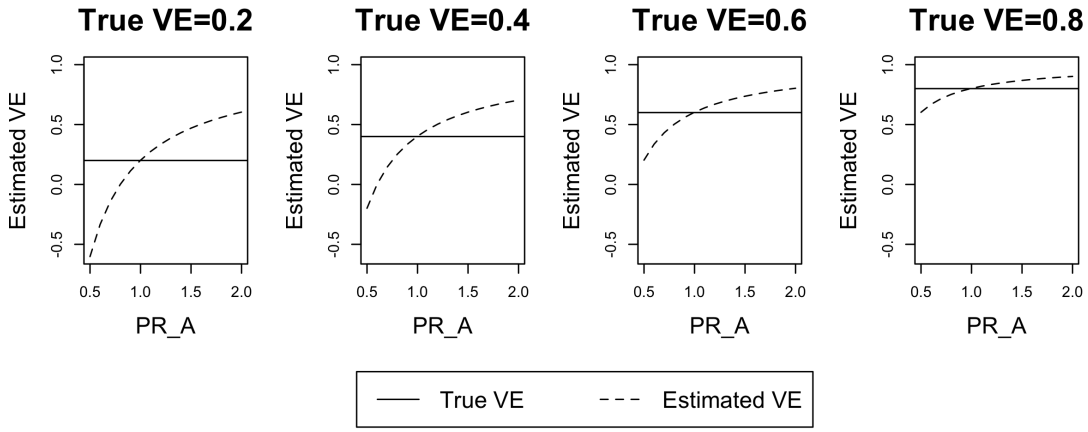
Each row of plots corresponds to a single source of bias and each column of plots corresponds to a single study design. Sources of bias: A - vaccination affects the probability of NFARI, BS - Healthy persons have a lower probability of NFARI and FARI, C - Vaccination lowers the probability of seeking medical care in FARI patients, and D - ARI patients with high health awareness have a higher probability of seeking medical care. Under sources of bias A, BS, and D, the true VE against SI is equal to the true VE against MAI. Under source of bias C the true VE against MAI differs from the true VE against SI when the probability ratio is not equal to 1. If only a solid line is visible, that indicates that the VE estimate is unbiased. Since we do not estimate VE against MAI from ASC studies, the figure for VE against MAI when source of bias C is present for ASC studies is blank. PR - probability ratio.

3.3.1 Sensitivity Analyses

Alternative Values of True VE

To assess whether the value of true VE had an effect on the magnitude of bias, we performed our calculations using different values of true VE (0.2, 0.4, 0.6, and 0.8). Table 3.11 shows the bias of VE estimates under select sources of bias for different values of true VE. Based on our results, bias is dependent on the value of the true VE (particularly for the case-control studies). When the true VE is higher, bias is lower. Much of the biases from A and C in the case-control studies are reduced when the true VE is 0.8; however, if true VE is 0.2, the bias is even larger than our original scenario (true VE=0.437). We illustrate this phenomenon in Figure 3.3 where the magnitude of bias under source of bias A in TN-based VE estimates shrinks as the true VE increases. Our results suggest that more effective vaccines are more robust to sources of bias.

Figure 3.3: VE estimates from TN studies under source of bias A for different values of true VE



PR_A=probability ratio under source of bias A (vaccination affects the probability of NFARI).

Table 3.11: Bias of estimates of VE in observational studies for different values of true VE

Source of Bias[a]	Outcome of Interest	True VE	ASC			PSC			TN			TCC		
			Q(5)[b]	Q(50)	Q(95)	Q(5)	Q(50)	Q(95)	Q(5)	Q(50)	Q(95)	Q(5)	Q(50)	Q(95)
A	SI & MAI	0.20	0.00	0.00	0.00	0.00	0.00	0.00	-0.48	0.01	0.30	-0.21	0.02	0.13
		0.40	0.00	0.00	0.00	0.00	0.00	0.00	-0.34	0.01	0.23	-0.15	0.02	0.1
		0.60	0.00	0.00	0.00	0.00	0.00	0.00	-0.24	0.00	0.15	-0.09	0.02	0.08
		0.80	0.00	0.00	0.00	0.00	0.00	0.00	-0.12	0.00	0.08	-0.04	0.01	0.04
BS	SI & MAI	0.20	-0.08	-0.02	0.00	-0.08	-0.02	0.00	0.00	0.00	0.00	-0.10	-0.02	0.01
		0.40	-0.06	-0.02	0.00	-0.06	-0.02	0.00	0.00	0.00	0.00	-0.06	-0.01	0.02
		0.60	-0.04	-0.01	0.00	-0.04	-0.01	0.00	0.00	0.00	0.00	-0.03	0.00	0.02
		0.80	-0.02	-0.01	0.00	-0.02	-0.01	0.00	0.00	0.00	0.00	-0.01	0.00	0.01
C	SI	0.20	0.00	0.00	0.00	0.01	0.14	0.33	0.02	0.15	0.33	0.03	0.15	0.34
		0.40	0.00	0.00	0.00	0.01	0.1	0.26	0.01	0.11	0.26	0.03	0.12	0.27
		0.60	0.00	0.00	0.00	0.01	0.07	0.16	0.01	0.08	0.16	0.03	0.09	0.17
		0.80	0.00	0.00	0.00	0.00	0.04	0.08	0.01	0.04	0.08	0.02	0.05	0.09
D	SI & MAI	0.20	0.00	0.00	0.00	-0.16	-0.05	0.00	0.00	0.00	0.00	-0.13	-0.04	0.01
		0.40	0.00	0.00	0.00	-0.11	-0.04	0.00	0.00	0.00	0.00	-0.08	-0.02	0.02
		0.60	0.00	0.00	0.00	-0.08	-0.03	0.00	0.00	0.00	0.00	-0.05	0.00	0.02
		0.80	0.00	0.00	0.00	-0.04	-0.01	0.00	0.00	0.00	0.00	-0.02	0.00	0.01
BS,D,A	SI & MAI	0.20	-0.07	-0.02	0.00	-0.2	-0.09	-0.02	-0.46	0.01	0.31	-0.41	-0.1	0.07
		0.40	-0.06	-0.02	0.00	-0.14	-0.06	-0.01	-0.34	-0.01	0.21	-0.25	-0.05	0.06
		0.60	-0.04	-0.01	0.00	-0.09	-0.05	-0.01	-0.25	-0.01	0.16	-0.19	-0.03	0.05
		0.80	-0.02	-0.01	0.00	-0.05	-0.02	0.00	-0.12	0.00	0.08	-0.08	-0.01	0.03
BS,D,A,C	SI	0.20	-0.07	-0.02	0.00	-0.11	0.08	0.28	-0.27	0.16	0.43	-0.20	0.07	0.30
		0.40	-0.06	-0.02	0.00	-0.09	0.05	0.21	-0.21	0.12	0.32	-0.17	0.07	0.24
		0.60	-0.04	-0.01	0.00	-0.06	0.04	0.14	-0.13	0.08	0.22	-0.10	0.05	0.16
		0.80	-0.02	-0.01	0.00	-0.03	0.02	0.07	-0.07	0.04	0.11	-0.05	0.03	0.08
BS,D,A,C	MAI[c]	varies[d]	–	–	–	-0.16	-0.07	-0.02	-0.36	0.00	0.24	-0.32	-0.07	0.06
			–	–	–	-0.13	-0.05	-0.01	-0.29	0.00	0.18	-0.23	-0.05	0.05
			–	–	–	-0.08	-0.03	-0.01	-0.18	0.01	0.13	-0.15	-0.02	0.04
			–	–	–	-0.04	-0.02	0.00	-0.10	0.01	0.06	-0.07	-0.01	0.02

[a] Source of bias: see Table 3.6. [b] Q(5)=5th quantile, Q(50)=50th quantile (median), and Q(95)=95th quantile. [c] We do not estimate VE against MAI for ASC studies. [d] True VE against MAI varies based on the value of the probability ratio when bias C was present.

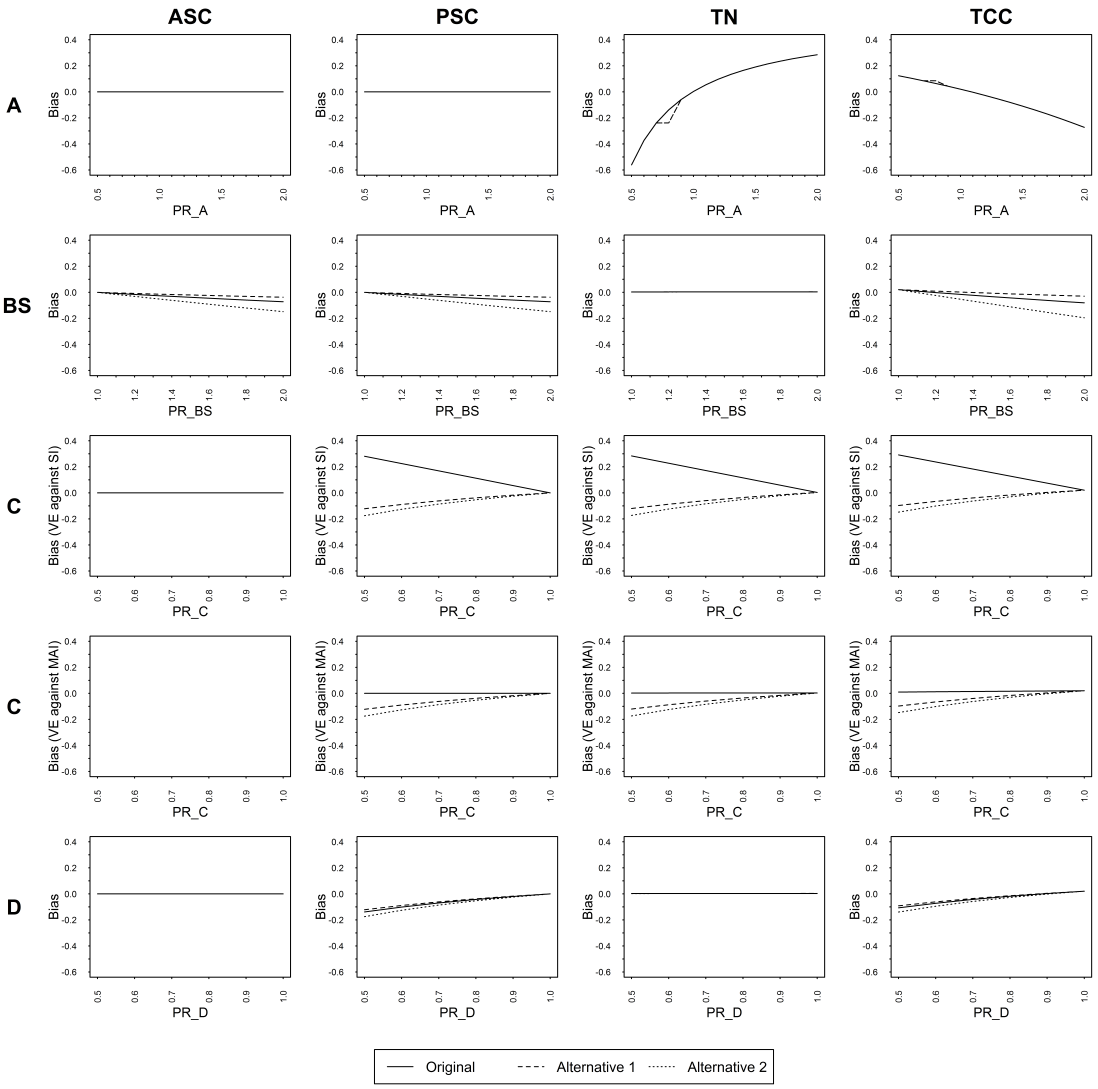
Alternative Probabilities of (X,U)

When A and C were present, bias was the same for both sets of alternative probabilities when compared to the original probabilities for all study designs. When B1 was present bias was the same as the original probabilities for the cohort studies, but differed for the case-control studies. When B2 and BS were present, bias using both sets of alternative probabilities differed from the original calculations for all study designs. When D was present, bias was the same as the original calculations for ASC and TN studies, but differed under both sets of alternative probabilities for PSC and TCC studies (Figure 3.4).

Alternative Probabilities of Vaccination

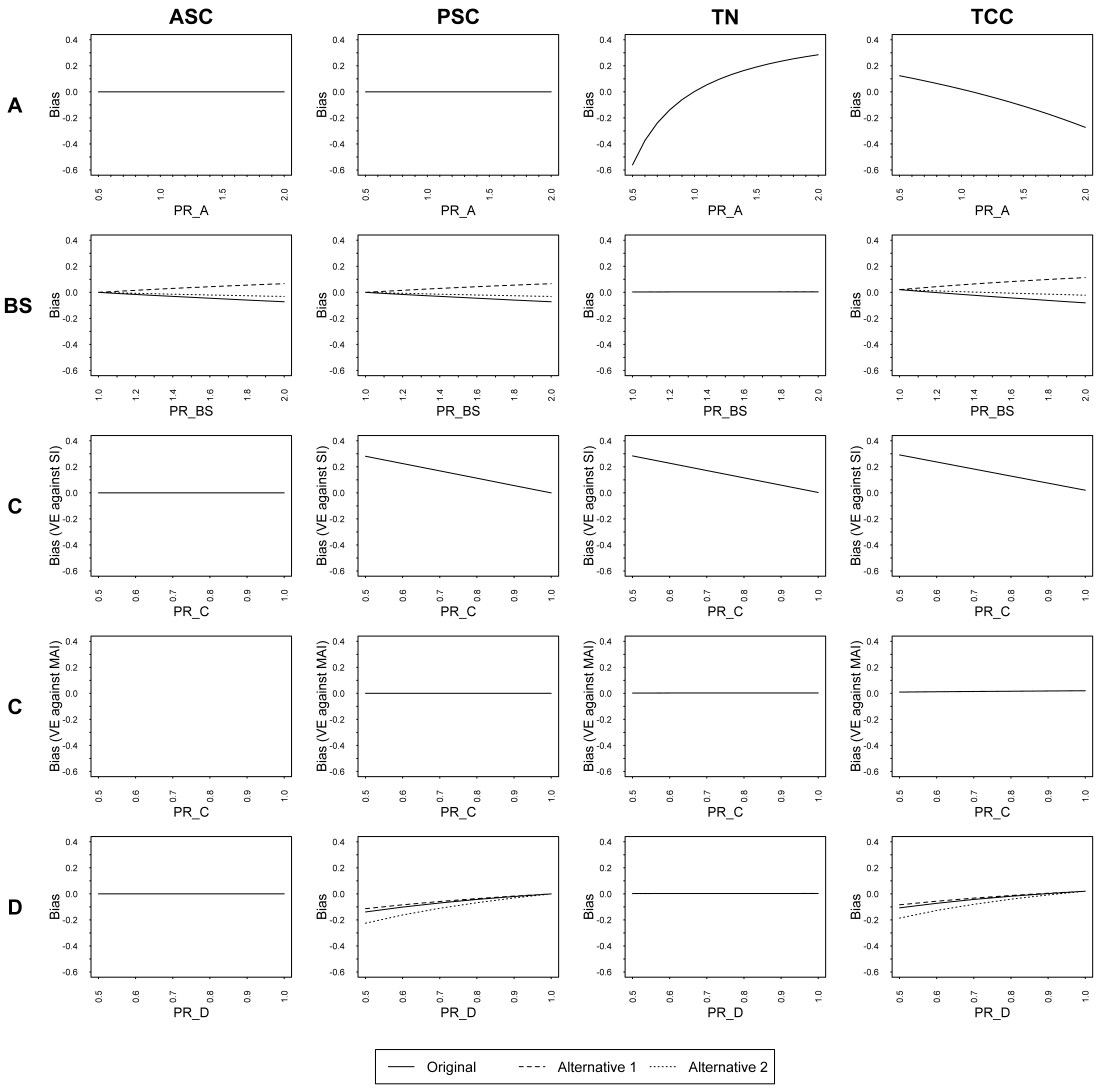
Bias was the same for each set of alternative input probabilities compared to the original input probabilities for all study designs when A and C were present. Bias of estimates of VE from the cohort studies was the same in the alternative scenarios compared to the original scenario, but differed for case-control studies when B1 was present. When B2 and BS were present, bias differed in the alternative scenarios compared to the original scenario for all study designs, except for bias in the TN study for alternative scenario 1. When D was present, bias was the same in the alternative scenarios compared to the original scenario for ASC and TN studies, but differed for all three alternative scenarios for PSC and TCC studies (Figure 3.5).

Figure 3.4: Bias of VE estimates from ASC, PSC, TN, and TCC studies for alternative probabilities of (X, U)



Each plot shows the bias of VE estimates for different probabilities of (X, U) for a single source of bias (row) and study design (column) as the probability ratio varies. Under sources of bias A, BS, and D, the true VE against SI is equal to the true VE against MAI thus, the bias is the same for both outcomes of interest. Under source of bias C the true VE against MAI differs from the true VE against SI when the probability ratio is not equal to 1. If only a solid line is visible the bias is the same among the different scenarios. Since we do not obtain an estimate of VE against MAI from ASC studies, the figure for VE against MAI when source of bias C is present for ASC studies is blank.

Figure 3.5: Bias of VE estimates from ASC, PSC, TN, and TCC studies for alternative probabilities of vaccination



Each plot shows the bias of VE estimates for different probabilities of vaccination for a single source of bias (row) and study design (column) as the probability ratio varies. Under sources of bias A, BS, and D, the true VE against SI is equal to the true VE against MAI thus, the bias is the same for both outcomes of interest. Under source of bias C the true VE against MAI differs from the true VE against SI when the probability ratio is not equal to 1. If only a solid line is visible the bias is the same among the different scenarios. Since we do not obtain an estimate of VE against MAI from ASC studies, the figure for VE against MAI when source of bias C is present for ASC studies is blank.

Sensitivity and Specificity

We varied the values of influenza test sensitivity and specificity to assess how deviations from our assumption of 100% sensitivity and specificity impacted influenza VE estimates. We assessed the median absolute value of bias for four combinations of sensitivity and specificity: (0.9, 0.9), (0.9, 1.0), (1.0, 0.9), and (1.0, 1.0). Our results suggest that VE estimates are robust to imperfect test sensitivity, but may be more biased when the test has less than perfect specificity (Table 3.12). However, when source of bias C is present and the outcome of interest is SI, estimates of VE from PSC, TN, and TCC have a smaller absolute value of bias when the specificity is 0.9. Additionally, when bias A is present estimates of VE from TN studies have similar median absolute value of bias regardless of the value of test sensitivity and specificity.

3.4 Discussion

In this chapter we present a dynamic extension of a previously developed model [99] for the evaluation of bias of influenza VE estimates from both case-control and cohort studies. This extended model provides several advantages over existing models [46, 55, 57, 98, 99], namely, 1) a time component allowing for the intensities of FARI and NFARI to change over time and the possibility of developing more than one ARI in a season, 2) the incorporation of two covariates (health status and health awareness) that may affect the probabilities of vaccination, developing ARIs, and seeking medical care for these ARIs, and 3) the ability to assess VE estimates from cohort studies. This work makes an important contribution to the literature by comparing the bias of VE estimates from four study designs under various sources of bias.

Table 3.12: Median absolute value of bias of estimated VE against SI and MAI for different values of influenza test sensitivity and specificity

Source of Bias[a]	Outcome of Interest	Sensitivity	Specificity							
			ASC		PSC		TN		TCC	
			0.9	1.0	0.9	1.0	0.9	1.0	0.9	1.0
A	SI and MAI	0.9	0.17	0.00	0.13	0.00	0.14	0.11	0.11	0.05
		1.0	0.16	0.00	0.12	0.00	0.12	0.12	0.10	0.05
BS	SI and MAI	0.9	0.20	0.02	0.15	0.02	0.13	0.00	0.13	0.02
		1.0	0.19	0.02	0.14	0.02	0.11	0.00	0.12	0.01
C	SI	0.9	0.18	0.00	0.06	0.10	0.06	0.10	0.05	0.12
		1.0	0.16	0.00	0.05	0.10	0.05	0.11	0.04	0.12
	MAI[b]	0.9	—	—	0.16	0.00	0.16	0.00	0.14	0.02
		1.0	—	—	0.15	0.00	0.14	0.00	0.12	0.02
D	SI and MAI	0.9	0.18	0.00	0.18	0.04	0.13	0.00	0.15	0.02
		1.0	0.16	0.00	0.16	0.04	0.12	0.00	0.13	0.02

[a] Sources of bias: A - vaccination affects the probability of NFARI, BS - Healthy persons have a lower probability of NFARI and FARI, C - Vaccination lowers the probability of seeking medical care in FARI patients (because of reduced symptom severity), and D - ARI patients with high health awareness have a higher probability of seeking medical care.[b] We do not estimate VE against MAI for ASC studies. Median absolute value of bias calculated from 1000 Monte Carlo simulations.

Our primary findings for each study design are:

- ASC studies produce unbiased estimates of VE against SI except when health status influences the probability of FARI (bias B2), in which case the bias may be moderate.
- PSC studies produce unbiased VE estimates against MAI except when health status influences the probability of FARI (bias B2) or health awareness influences the probability of seeking care (bias D). This study design may produce severely biased estimates against SI when vaccination lowers the probability of seeking medical care.
- VE estimates from TN studies are unbiased for both SI and MAI when health status has the same effect on the probabilities of FARI and NFARI (bias BS) and when health awareness affects the probability of seeking care (bias D). Estimates are unbiased for MAI when vaccination lowers the probability of seeking care for FARI (bias C). TN-based VE estimates may be severely biased when vaccination affects the probability of NFARI (bias A) for both outcomes or when vaccination affects the probability of seeking care (bias C) and the outcome of interest is SI.
- TCC studies may produce moderate to severely biased VE estimates in the presence of all sources of bias.

Our results confirm earlier findings [55, 99] that when vaccination affects the probability of NFARI (bias A), TN-based estimates may be severely biased. We also show that under this source of bias VE estimates from TCC studies may be severely biased, but VE estimates from both cohort studies are unbiased.

In the results presented above, we only considered one value of true VE and one possible combination of the probabilities of health status, health awareness,

and vaccination. To assess the impact of the value of true VE on the magnitude of bias, we varied the value of true VE. Based on our results, bias is dependent on the value of the true VE, and estimates of more effective vaccines are less influenced by the sources of bias evaluated in this work compared to less effective vaccines. Additionally, we assumed that health status and health awareness were independent. To assess the impact of different input probabilities on the bias of VE estimates from each study design, we varied these probabilities. We observed no change in the bias of VE estimates from any study design when biases A and D were present. Bias did differ when the other sources of bias were present. Especially, VE estimates from the TN were no longer unbiased against MAI in the presence of source of bias C when the probabilities of health status and health awareness were correlated. Finally, we varied the value of influenza test sensitivity and specificity and observed that influenza VE estimates are robust to imperfect test sensitivity, but in general, had a larger magnitude of median absolute value bias when the influenza test had imperfect specificity.

It is important to consider that the biases presented here are from unadjusted VE estimates. We focused on characterizing bias from different sources, rather than on methods to adjust for this bias. It is reasonable to postulate that if appropriate adjustments are made, some of the bias may be reduced. Another limitation of this work is that we did not account for the misclassification of vaccination status.

Based on the results presented here, and the assumption that multiple sources of bias are present, the preferred study designs for estimating VE against SI and MAI are ASC and TN, respectively. TN studies are cheaper and involve fewer logistical issues compared to ASC studies. The TN is the only study design that adjusts for the two most common sources of bias: BS and D, which represent

the confounding due to covariates (health status and health awareness) that are related to both vaccination and the outcome. However, if the core assumption underlying this study design is violated, i.e., bias A is present, then one should consider a cohort study, which is usually more expensive and logistically complex. Even if the core assumption is satisfied, one should not interpret TN-based estimates as estimates of VE against SI, unless bias C has been ruled out.

Chapter 4

Estimation of bias of influenza vaccine effectiveness estimates from test-negative studies: extensions of a dynamic probability model

Since first being used to assess influenza VE in 2005 [56], the TN study design has become the most popular design for assessing annual influenza VE. The TN design is attractive because it can be easily incorporated into existing surveillance systems and attempts to control for confounding due to propensity to seek medical care because cases and controls are both selected from individuals who seek medical care for ARI [100]. Within a TN study, cases are selected from individuals who seek medical care for ARI and test positive for influenza infection, while controls are individuals who seek care for ARI and test negative for influenza infection.

Concern has been raised with the growing popularity of the TN design about the validity of TN-based VE estimates. Previous work has demonstrated

the validity of TN-based estimates of VE if vaccination does not affect the probability of developing NFARI in outpatient settings [45, 55, 57, 98, 106], in inpatient settings [32], and when the influenza test has imperfect sensitivity and specificity [46]. However, more work is needed to determine the validity of TN-based estimates under scenarios that deviate from commonly made assumptions. In this chapter, we present two extensions of the model presented in Chapter 3.

First, we consider the scenario when vaccination occurs during the study period rather than assume all vaccinated individuals were vaccinated prior to the study period, as has been done previously [32, 45, 46, 55, 57, 98, 99, 106, 124]. The assumption of vaccination prior to the study period is reasonable for seasonal epidemics because vaccine campaigns begin prior to the outbreak; however, it is not a reasonable assumption during an influenza pandemic. For example, during the 2009 influenza pandemic, a vaccine was made available months after the start of the outbreak [7]. We extend our model from Chapter 3, to allow vaccination to occur over time and assess the bias of TN-based estimates of VE under different sources of bias. We then compare the bias of estimates from our extended model to the bias of estimates when vaccination occurs prior to the study period.

Second, we assume a different vaccine model, the all-or-none vaccine model [44]. In Chapter 3 we assumed a leaky vaccine model, where vaccination lowers the probability of infection a fraction, ρ. In the all-or-none model, a proportion ρ of vaccinated individuals are conferred complete immunity from infection, while the remaining proportion $(1 - \rho)$ are fully susceptible to infection. We assess the bias of TN-based estimates under the all-or-none model under different sources of bias and compare them to TN-based estimates under the leaky

model.

For each extension, we make the same assumptions detailed in section 3.2.1 with the additional assumption that the influenza test has perfect sensitivity and specificity.

4.1 Vaccination Over Time

4.1.1 Background

Since the beginning of the 20th century, five influenza pandemics have occurred, none more devastating than the 1918 'Spanish flu' pandemic. Despite the medical advances (e.g., vaccines and antivirals) and increased public awareness throughout the previous century, influenza pandemics pose a significant public health threat. The most recent 2009 H1N1 pandemic provided a powerful reminder of how dangerous pandemic influenza can be. An estimated 151,700 to 575,400 deaths occurred worldwide during the first year the pandemic strain circulated [125]. Due to the continual adaptation of the influenza A virus, future influenza pandemics are inevitable. One important element of pandemic preparedness is the rapid development of a vaccine against the pandemic strain, as vaccination remains the best way to prevent against influenza infection [7]. It is estimated that as many as 1.5 million cases, 4,000-10,000 hospitalizations and 200-500 deaths were averted in the United States by the monovalent vaccine during the 2009 pandemic [10].

In the context of a pandemic, estimation of influenza vaccine effectiveness (VE) involves additional challenges compared to the estimation of the effectiveness of seasonal influenza vaccines. During the 2009 pandemic, the monovalent vaccine against the pandemic strain was made available months after the start

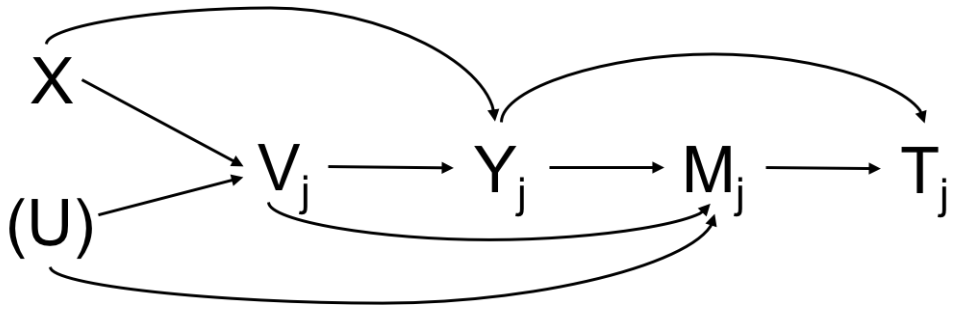
of the pandemic. The delayed and gradual timing of vaccination may introduce additional bias into estimates of VE compared to seasonal epidemics, where most people get vaccinated before the outbreak.

4.1.2 Methods

Model Description

We present an extension to the dynamic model presented in Chapter 3 that allows vaccination status to change over time (Figure 4.1). Details about the associated variables and the probabilities determining each variable's distribution can be found in Tables 4.1 and 4.2, respectively. All variables are defined for each member of the study population. Some variables vary over time, and we consider each time unit to be one week. It will be convenient to define a *standard person* as a person who is unvaccinated, healthy, and has high health awareness. In Table 4.2, probabilities are first defined for a standard person, and then defined for non-standard persons as a function of the corresponding *standard person* probability and multiplier(s).

Figure 4.1: Causal graph of influenza vaccine studies with covariates.



X = health status, (U) = health awareness (unobserved), V_j = vaccination status in week j, Y_j = ARI status in week j, M_j = seeking medical care for ARI in week j, and T_j = influenza test result in week j, where $j = 1, \ldots, J$ and J = the number of weeks in the study.

Step 1: Covariates. We assume that people within the population can be classified with a health status (X) of either "healthy" or "frail" and a health awareness (U) of either "high" or "low".

Step 2: Vaccination. Individuals can be vaccinated throughout the study. We use j_v to denote the first week individuals may become vaccinated. A person is considered effectively vaccinated 14 days after receipt of the vaccine. We will use the term 'vaccinated' to indicate 'effectively vaccinated'.

Step 3: Influenza and non-influenza ARI. During the study, a person may become infected with an influenza virus and develop FARI and/or develop one or more NFARIs. We define a variable Y_j for the illness/infection status in week j, where the distribution of Y_j may depend on the person's vaccination status (V_j) and health status (X).

Step 4: Seeking medical care for ARI. A person with an ARI in week j may seek medical care (M_j). The probability of seeking medical care depends on Y_j, as only those individuals who have an ARI may seek medical care (i.e., $M_j = 0$

if $Y_j = 0$), and it may be different for FARI and NFARI patients. This probability may also depend on V_j and U.

Step 5: Testing for influenza infection. We assume that each person who seeks medical care for ARI is tested for influenza infection (T_j).

Table 4.1: Variables in the model.

Variable	Definition	Values
X	Health status	0 - frail person 1 - healthy person
U	Health awareness (unobserved)	0 - low health awareness 1 - high health awareness
V_j	Vaccination status in week j	0 - unvaccinated 1 - vaccinated
K	The week a person became effectively vaccinated	$K = j_v, \ldots, J+1$*
Y_j	Influenza/non-influenza ARI status in week j	0 - no ARI 1 - non-influenza ARI 2 - influenza ARI
M_j	Seeking medical care for ARI in week j	0 - no 1 - yes
T_j	Result of test for influenza infection in week j	0 - negative 1 - positive

*$K = j_v$ indicates the first week individuals may become vaccinated and $K = J + 1$ indicates a person who was not vaccinated by the end of the study (i.e., remain unvaccinated).

True VE

The true VEs against SI and MAI may be different [99, 124]. Therefore, we evaluated the true VE for each of the two outcomes of interest. True VE is calculated under the assumption of random vaccination, i.e., the probability that $K = j$ does not depend on X or U. Because vaccination status can change during the study, we derive true VE allowing persons to be vaccinated during the study. The true VE is one minus the ratio of $\frac{\sum_j C_{jv}}{\sum_j N_{jv}}$ when $v = 1$ and $v = 0$,

Table 4.2: Parameters in the model.

Parameters	Definition	Comments
π_{xu}	$\mathbb{P}(X = x, U = u)$	
α_{jxu}	$\mathbb{P}(K = j \mid X = x, U = u)$	
β_{jvx}	$\mathbb{P}(Y_j = 1 \mid V_j = v, X = x)$	$\beta_{j11} = \beta_{j01} \cdot \theta_\beta$
θ_β	multiplier for β when $V_j = 1$	$\beta_{j00} = \beta_{j01} \cdot \phi_\beta$
ϕ_β	multiplier for β when $X = 0$	$\beta_{j10} = \beta_{j01} \cdot \theta_\beta \cdot \phi_\beta$
γ_{jvx}	$\mathbb{P}(Y_j = 2 \mid V_j = v, X = x)$	$\gamma_{j11} = \gamma_{j01} \cdot \theta_\gamma$
θ_γ	multiplier for γ when $V_j = 1$	$\gamma_{j00} = \gamma_{j01} \cdot \phi_\gamma$
ϕ_γ	multiplier for γ when $X = 0$	$\gamma_{j10} = \gamma_{j01} \cdot \theta_\gamma \cdot \phi_\gamma$
δ_{1vu}	$\mathbb{P}(M_j = 1 \mid Y_j = 1, V_j = v, U = u)$	$\delta_{111} = \delta_{101} \cdot \theta_{\delta_1}$
θ_{δ_1}	multiplier for δ_1 when $V_j = 1$	$\delta_{100} = \delta_{101} \cdot \mu_{\delta_1}$
μ_{δ_1}	multiplier for δ_1 when $U = 0$	$\delta_{110} = \delta_{101} \cdot \theta_{\delta_1} \cdot \mu_{\delta_1}$
δ_{2vu}	$\mathbb{P}(M_j = 1 \mid Y_j = 2, V_j = v, U = u)$	$\delta_{211} = \delta_{201} \cdot \theta_{\delta_2}$
θ_{δ_2}	multiplier for δ_2 when $V_j = 1$	$\delta_{200} = \delta_{201} \cdot \mu_{\delta_2}$
μ_{δ_2}	multiplier for δ_2 when $U = 0$	$\delta_{210} = \delta_{201} \cdot \theta_{\delta_2} \cdot \mu_{\delta_2}$

A *standard person* is defined as a person with ($V_j = 0, X = 1, U = 1$). β_{j01} and γ_{j01} represent the probabilities of contracting a NFARI or FARI for a *standard person* in week j, respectively, $j = 1, \ldots, J$. δ_{101} and δ_{201} represent the probabilities of seeking care for NFARI and FARI for a *standard person*, respectively. These probabilities, as well as all π_{xu}, $x = 0, 1; u = 0, 1$, all α_{jxu}, $j = 1, \ldots, J$; $x = 0, 1$; $u = 0, 1$, τ_1, τ_2, and all multipliers ($\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$), are input parameters.

where C_{jv} and N_{jv} are the expected numbers of cases and persons of vaccine status v in week j, respectively. A person is considered a true case of SI if s/he develops an influenza ARI during the study ($Y_j = 2$). A person is considered a true case of MAI if s/he develops an influenza ARI during the study and seeks medical care for this ARI ($Y_j = 2, M_j = 1$). See Appendix C.1.1 for expressions of true VE in terms of model parameters.

VE Estimates from TN Studies

We assume that a person is classified as a TN case or a TN control at her/his first ARI-related visit. This classification does not change, regardless of possible conflicting test results in future visits. Similarly, when a person is classified as a case or control, only their vaccination status at that visit is recorded, regardless of whether their vaccination changes in future visits. We assume that the study begins in week $j = j_v$, the first week of vaccination. Therefore, we define $M_j = 0$ for $j < j_v$.

Cases

A person is considered a case in week j, if:

- they did not seek medical care for any ARI prior to week j, so $M_k = 0$ for every week $k = 1, \ldots, j-1$ (i.e., $\mathbf{M_{j-1} = 0}$)
- they seek medical care for their ARI in week j (i.e., $M_j = 1$)
- they have FARI in week j (i.e., $Y_j = 2$)

The expected number of vaccinated cases in week j is:

$$\mathbb{E}(\text{unvaccinated cases in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1} = 0}, M_j = 1, Y_j = 2, K \leq j)$$

and the expected number of unvaccinated cases in week j is:

$$\mathbb{E}(\text{unvaccinated cases in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1} = 0}, M_j = 1, Y_j = 2, K > j)$$

See Appendix C.1.2 for expressions of expected numbers of cases in terms of the model parameters.

Controls

A control in week j is defined in the same way as a case with the exception that $Y_j = 1$. The expected number of vaccinated controls in week j is:

$$\mathbb{E}(\text{unvaccinated controls in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K \leq j)$$

and the expected number of unvaccinated controls in week j is:

$$\mathbb{E}(\text{unvaccinated controls in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K > j)$$

For expressions of expected number of controls in terms of the model parameters, see Appendix C.1.2. To obtain the final 2x2 table for a TN study (Table 4.3), we sum the expected cell counts over all weeks in the study period $j_v, \ldots, J$. The VE estimate from a TN study is

$$\hat{VE}_{TN} = 1 - \hat{OR}_{TN},$$

where $\hat{OR}_{TN}$ is the odds ratio from Table 4.3.

Table 4.3: Final 2x2 table for a TN study.

	Vaccinated	**Unvaccinated**
Case	$\sum_{j=j_v}^{J} \mathbb{E}(\text{vaccinated cases in week } j)$	$\sum_{j=j_v}^{J} \mathbb{E}(\text{unvaccinated cases in week } j)$
Control	$\sum_{j=j_v}^{J} \mathbb{E}(\text{vaccinated controls in week } j)$	$\sum_{j=j_v}^{J} \mathbb{E}(\text{unvaccinated controls in week } j)$

Calculations and Simulations

Using expressions of true and estimated VE derived from our model (Appendices C.1.1 and C.1.2), we calculated the bias of TN-based VE estimates under various sources of bias (Table 4.4). Bias was defined as estimated VE minus true VE. Each source of bias can be attributed to deviation of a specific parameter from 1.0, and bias was calculated by varying the values of the corresponding parameters (Table 4.4). When a source of bias was absent, we kept the corresponding parameter fixed at 1.0. We calculated VE estimates from a 20 week study ($J = 20$) in which the first week of vaccination varied ($j_v = 1, 6, 11, 15, 20, 24, 28, 33, 38$). The values of j_v corresponded to the first week of each month during the 2009 influenza pandemic season (with the exception of week 1, which corresponded to the beginning of the outbreak). For bias A,

Table 4.4: Sources of Bias

Label	Source of Bias	Parameter	Range
A	Vaccination affects the probability of NFARI	θ_β	0.5-2.0
B1	Healthy persons have a lower probability of NFARI	ϕ_β	1.0-2.0
B2	Healthy persons have a lower probability of FARI	ϕ_γ	1.0-2.0
BS	Healthy persons have a lower probability of FARI and NFARI. Health status has the same effect on the probabilities of both types of ARI.	$\phi_\beta = \phi_\gamma$	1.0-2.0
C	Vaccination lowers the probability of seeking medical care in FARI patients (because of reduced symptom severity).	θ_{δ_2}	0.5-1.0
D	ARI patients with high health awareness have a higher probability of seeking medical care.	$\mu_{\delta_1} = \mu_{\delta_2}$	0.5-1.0

we considered that vaccination might increase or decrease the probability of NFARI, so θ_β, the ratio of the probability of NFARI in vaccinated persons compared to unvaccinated persons, varied from 0.5 to 2.0. For biases B1, B2, and BS, we allowed ϕ_β and ϕ_γ, the ratios of the probabilities of NFARI and FARI in frail

persons compared to healthy persons, respectively, to vary between 0.5 and 1.0, since we expect healthy persons to have lower probabilities of ARI compared to frail persons. For bias C, θ_{δ_2}, the ratio of the probability of seeking medical care for FARI in vaccinated persons compared to unvaccinated persons, varied between 0.5 to 1.0, since we expect vaccination to reduce the probability of seeking medical care for FARI compared to NFARI. For bias D, we let μ_{δ_1} and μ_{δ_2} denote the ratios of the probability of seeking medical care for NFARI and FARI, respectively, comparing persons with high health awareness to persons with low health awareness. We let their values vary between 0.5 to 1.0 because we expect persons with high health awareness to have a higher probability of seeking medical care for both FARI and NFARI compared to persons with low health awareness. We assume $\mu_{\delta_1} = \mu_{\delta_2}$.

Under each source of bias we determined the 5^{th}, 50^{th}, and 95^{th} quantiles of bias and the 50^{th} and 95^{th} quantiles of the absolute value of bias from 1,000 Monte Carlo simulations. For each simulation, values for the relevant parameter(s) were drawn from independent triangular distributions over the ranges specified in Table 4.4. The mode of each distribution was assumed to be 1. The true and estimated VE were calculated for each simulation. Bias was calculated for each source of bias separately as well as for combinations of sources of bias. We validated the calculations using a stochastic simulation program (Appendix C.1.3). The mean bias of VE estimates over 1,000 simulations were compared to the calculated values. All simulations were performed in R 3.3.1 [112].

Sensitivity Analyses

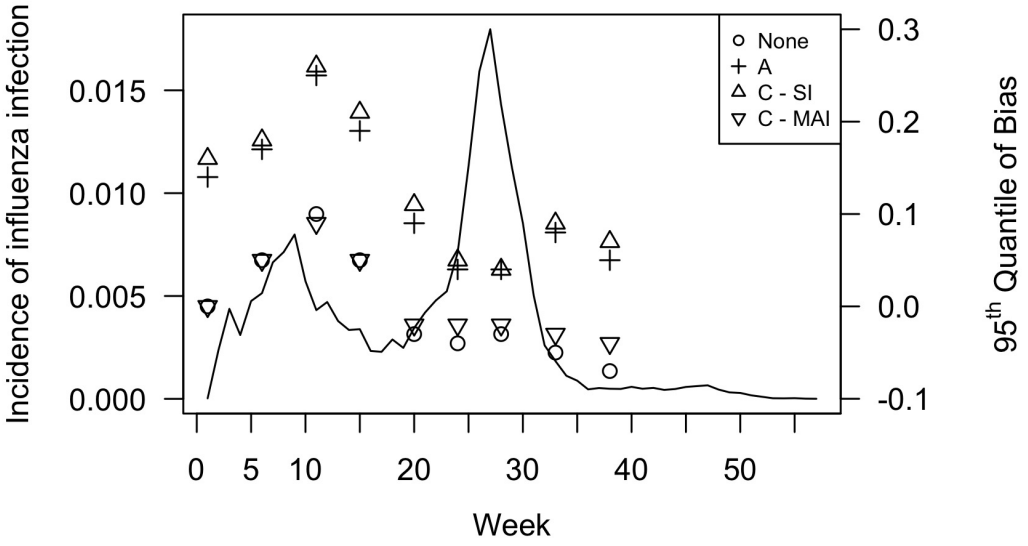
We performed several sensitivity analyses to determine if different parameter values affected the magnitude of bias of under the all-or-none vaccine model. First, we varied vaccination coverage (0.40, 0.60, and 0.80). Next, we varied the value of true VE (0.2, 0.4, 0.6, and 0.8). Finally, we varied the length of the study period (15, 20, 25, 30 weeks). For each sensitivity analysis we held the first week of vaccination fixed at week 24.

4.1.3 Results

We evaluated bias of VE estimates from TN studies in the presence of the sources of bias listed in Table 4.4. We calculated the 5th, 50th, and 95th quantiles of bias and the 50th and 95th quantiles of the absolute value of bias under each source of bias (Table 4.5). We define several terms to aid in our evaluation of the magnitude of bias: little/small or no bias indicates an absolute bias of less than 0.05, moderate bias indicates absolute bias greater than or equal to 0.05 and less than 0.10, substantial bias indicates absolute bias greater than or equal to 0.10 and less than 0.20, and severe bias indicates absolute bias of 0.20 or more.

When vaccination occurred during the study, the effect of the first week of vaccination (j_v) was unpredictable. Interestingly, the 95th quantile of bias of VE estimates from different sources followed a similar pattern when the first week of vaccination varied (Figure 4.2). The TN study produced biased VE estimates regardless of the source of bias present, except when all vaccinated individuals were vaccinated in week 1 (baseline). The values of bias under biases BS and D were the same as when no sources of bias were present (Table 4.5). The largest bias under these sources of bias was observed at $j_v = 15$ (0.10) compared to

Figure 4.2: 95th Quantile of absolute value of bias of VE estimates by first week of vaccination



The solid line indicates the incidence of influenza infection in each week. The 95th quantile of bias was the same for no sources of bias (None) and biases BS and D; therefore, BS and D are not represented in the figure individually. The 95th quantile of bias follows a similar pattern regardless of the source of bias as the first week of vaccination varies.

unbiased estimates at baseline. When vaccination influenced the probability of NFARI (bias A), VE estimates suffered from substantial to severe bias regardless of the value of j_v. The largest 95th quantiles of absolute value of bias were observed at $j_v = 11$ and $j_v = 38$ ($Q_{AVB}(95) = 0.26$); however this was only slightly larger than at baseline ($Q_{AVB}(95) = 0.24$). For several values of j_v, the bias of VE estimates was smaller than baseline. The smallest bias was observed when $j_v = 24$ ($Q_{AVB}(95) = 0.13$). When health status affected the probability of NFARI (bias B1), VE estimates suffered from little bias at baseline

($Q_{AVB}(95) = 0.03$) and little to substantial bias for all other values of j_v. The largest bias was observed at $j_v = 11$ ($Q_{AVB}(95) = 0.12$). When health status affected the probability of FARI (bias B2), VE estimates suffered from little bias at baseline ($Q_{AVB}(95) = 0.02$) and little to moderate bias for all other values of j_v, except $j_v = 38$, which suffered from substantial bias ($Q_{AVB}(95) = 0.10$). When vaccination lowers the probability of seeking medical care for FARI (bias C) the bias of VE estimates differed by outcome of interest. TN-based estimates of VE against SI suffered from little to severe bias depending on the first week of vaccination. At baseline, VE estimates were substantially biased ($Q_{AVB}(95) = 0.16$). The largest bias was observed at $j_v = 11$ ($Q_{AVB}(95) = 0.26$) and the smallest bias was observed at $j_v = 28$ ($Q_{AVB}(95) = 0.04$). When the outcome of interest was MAI, VE estimates at baseline were unbiased. For all other values of j_v VE estimates suffered from little to moderate bias ($Q_{AVB}(95) = 0.09$ at $j_v = 11$ and $Q_{AVB}(95) = 0.02$ at $j_v = 28$). We also assessed the bias of VE estimates under combinations of sources of bias (Table C2). Similar patterns of bias were observed under combinations of sources of bias as under a single source of bias.

Table 4.5: Bias of TN-based estimates of VE when first week of vaccination varies

Source of Bias[a]	**Outcome of Interest**	j_v	$Q_B(5)$[b]	$Q_B(50)$	$Q_B(95)$	$Q_{AVB}(50)$	$Q_{AVB}(95)$
None	SI & MAI	1 [c]	0.00	0.00	0.00	0.00	0.00
		6	0.05	0.05	0.05	0.05	0.05
		11	0.10	0.10	0.10	0.10	0.10
		15	0.05	0.05	0.05	0.05	0.05
		20	-0.03	-0.03	-0.03	0.03	0.03
		24	-0.04	-0.04	-0.04	0.04	0.04
		28	-0.03	-0.03	-0.03	0.03	0.03
		33	-0.05	-0.05	-0.05	0.05	0.05
		38	-0.07	-0.07	-0.07	0.07	0.07
A	SI & MAI	1	-0.24	0.00	0.14	0.08	0.24
		6	-0.14	0.05	0.17	0.08	0.18
		11	-0.16	0.09	0.25	0.12	0.26
		15	-0.18	0.04	0.19	0.09	0.22
		20	-0.24	-0.03	0.09	0.06	0.24
		24	-0.16	-0.04	0.04	0.05	0.16
		28	-0.13	-0.03	0.04	0.04	0.13
		33	-0.25	-0.05	0.08	0.07	0.25
		38	-0.26	-0.07	0.05	0.08	0.26
B1	SI & MAI	1	0.00	0.01	0.03	0.01	0.03
		6	0.05	0.06	0.07	0.06	0.07
		11	0.10	0.10	0.12	0.10	0.12
		15	0.05	0.06	0.07	0.06	0.07
		20	-0.03	-0.02	-0.01	0.02	0.03
		24	-0.04	-0.04	-0.03	0.04	0.04
		28	-0.03	-0.03	-0.02	0.03	0.03
		33	-0.04	-0.04	-0.03	0.04	0.04
		38	-0.07	-0.07	-0.06	0.07	0.07
B2	SI & MAI	1	-0.02	-0.01	0.00	0.01	0.02
		6	0.03	0.04	0.05	0.04	0.05
		11	0.07	0.09	0.09	0.09	0.09
		15	0.02	0.04	0.05	0.04	0.05
		20	-0.05	-0.04	-0.03	0.04	0.05
		24	-0.05	-0.05	-0.04	0.05	0.05
		28	-0.04	-0.03	-0.03	0.03	0.04
		33	-0.06	-0.05	-0.05	0.05	0.06
		38	-0.10	-0.08	-0.08	0.08	0.10

[a] Sources of bias: see Table 4.4, [b] $Q_B(5)$=5th quantile, $Q_B(50)$=50th quantile (median), $Q_B(95)$=95th quantile, $Q_{AVB}(50)$=50th quantile of the absolute value of bias, $Q_{AVB}(95)$=95th quantile of the absolute value of bias. Quantiles were determined from 1000 Monte Carlo simulations. Absolute value of bias is defined as the difference between the estimate VE and the true VE without regard to the sign, [c] all vaccinated individuals are vaccinated in the first week of the study.

Table 4.5: Bias of TN-based estimates of VE when first week of vaccination varies

Source of Bias	Outcome of Interest	j_v	$Q_B(5)$	$Q_B(50)$	$Q_B(95)$	$Q_{AVB}(50)$	$Q_{AVB}(95)$
BS	SI & MAI	1	0.00	0.00	0.00	0.00	0.00
		6	0.05	0.05	0.05	0.05	0.05
		11	0.10	0.10	0.10	0.10	0.10
		15	0.05	0.05	0.05	0.05	0.05
		20	-0.03	-0.03	-0.03	0.03	0.03
		24	-0.04	-0.04	-0.04	0.04	0.04
		28	-0.03	-0.03	-0.03	0.03	0.03
		33	-0.05	-0.05	-0.05	0.05	0.05
		38	-0.08	-0.08	-0.07	0.08	0.08
C[d]	SI	1	0.01	0.07	0.16	0.07	0.16
		6	0.05	0.11	0.18	0.11	0.18
		11	0.10	0.17	0.26	0.17	0.26
		15	0.06	0.12	0.21	0.12	0.21
		20	-0.02	0.03	0.11	0.03	0.11
		24	-0.04	0.00	0.05	0.02	0.05
		28	-0.03	0.00	0.04	0.02	0.04
		33	-0.04	0.02	0.09	0.03	0.09
		38	-0.07	-0.01	0.07	0.04	0.07
	MAI	1	0.00	0.00	0.00	0.00	0.00
		6	0.03	0.04	0.05	0.04	0.05
		11	0.06	0.08	0.09	0.08	0.09
		15	0.03	0.04	0.05	0.04	0.05
		20	-0.03	-0.02	-0.02	0.02	0.03
		24	-0.04	-0.03	-0.02	0.03	0.04
		28	-0.03	-0.02	-0.02	0.02	0.02
		33	-0.04	-0.04	-0.03	0.04	0.04
		38	-0.07	-0.06	-0.04	0.06	0.07
D	SI & MAI	1	0.00	0.00	0.00	0.00	0.00
		6	0.05	0.05	0.05	0.05	0.05
		11	0.10	0.10	0.10	0.10	0.10
		15	0.05	0.05	0.05	0.05	0.05
		20	-0.03	-0.03	-0.03	0.03	0.03
		24	-0.04	-0.04	-0.04	0.04	0.04
		28	-0.03	-0.03	-0.03	0.03	0.03
		33	-0.05	-0.05	-0.05	0.05	0.05
		38	-0.08	-0.08	-0.07	0.08	0.08

[d] Under bias C, estimates of VE against SI and MAI differ.

4.1.4 Sensitivity Analyses

We performed several sensitivity analyses to determine if different parameter values affected the magnitude of bias when vaccination started at different weeks during the season. For all sensitivity analyses, we compared the 95th quantile of absolute value of bias when vaccination occured prior to the season ($j_v = 1$) and at weeks 24 and 38. We varied the final vaccination coverage (0.20, 0.40, 0.60, and 0.80), value of true VE (0.2, 0.4, 0.6, and 0.8), and length of the study period (15, 20, 25, 30 weeks).

Vaccination Coverage

To assess whether the final vaccination coverage at the end of the study had an effect on the magnitude of bias when vaccination occured during the study period, we calculated the bias of TN-based VE estimates when the final vaccination coverage was 0.20, 0.40, 0.60, and 0.80. Under most sources of bias (except A and C against SI), the 95th quantile of absolute value of bias of VE estimates was larger when vaccination started in weeks 24 and 38 compared to baseline (Table 4.6). The largest difference in bias relative to baseline was 0.1 (under no bias, bias BS, or D when $j_v = 38$ and final vaccination coverage was 0.8). Under bias A, the bias was smaller when $j_v = 24$ compared to baseline. When $j_v = 38$, the bias was slightly larger than baseline. Under bias C when the outcome of interest was SI, the bias was largest at baseline ($Q_{AVB}(95) = 0.16$) and smallest when $j_v = 24$ ($Q_{AVB}(95)$ ranged from 0.04 to 0.06). Final vaccination coverage had little impact on the bias of VE estimates for a given value for j_v.

Table 4.6: Bias of TN-based estimates of VE when vaccination coverage varies

Source of Bias[a]	Outcome of Interest	j_v	Vaccination Coverage[b] 0.20	0.40	0.60	0.80
None	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.04	0.04	0.05	0.05
		38	0.07	0.08	0.09	0.10
A	SI & MAI	1	0.24	0.24	0.23	0.24
		24	0.17	0.17	0.16	0.17
		38	0.28	0.27	0.28	0.30
B1	SI & MAI	1	0.02	0.03	0.04	0.09
		24	0.04	0.04	0.05	0.05
		38	0.07	0.08	0.09	0.10
B2	SI & MAI	1	0.02	0.03	0.05	0.11
		24	0.05	0.06	0.06	0.07
		38	0.09	0.10	0.12	0.13
BS	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.04	0.04	0.05	0.05
		38	0.07	0.08	0.09	0.10
C[c]	SI	1	0.16	0.16	0.16	0.16
		24	0.06	0.05	0.04	0.05
		38	0.07	0.08	0.08	0.09
	MAI	1	0.00	0.00	0.00	0.00
		24	0.04	0.04	0.04	0.05
		38	0.07	0.08	0.09	0.10
D	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.04	0.04	0.05	0.05
		38	0.07	0.08	0.09	0.10

[a] For definitions of sources of bias, see Table 4.4. [b] Vaccination coverage refers to the final vaccination coverage at the end of the study. [c] Bias for estimates of VE against SI and MAI differ when bias C is present. The 95th quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, j_v, and vaccination coverage.

True VE

We varied the value of true VE (0.2, 0.4, 0.6, 0.8) to assess whether the value of true VE had an impact on the bias of estimates when vaccination occurred during the study period compared to baseline. Regardless of the value of true

VE, we observed a similar pattern of bias as in the original simulations. Under most sources of bias (except A and C against SI), the 95th quantile of absolute value of bias of VE estimates was larger when vaccination started in weeks 24 and 38 compared to baseline (Table 4.7). The largest difference in bias relative to baseline was 0.16 (under no bias, bias BS, or D when $j_v = 38$ and the true VE was 0.2). Under bias A, the bias was smaller when $j_v = 24$ compared to baseline. When $j_v = 38$, the bias was larger than baseline. For all values of true VE and j_v estimates suffered from substantial to severe bias. Under bias C when the outcome of interest was SI, the bias was largest at baseline ($Q_{AVB}(95)$ ranged from 0.08 to 0.34) and smallest when $j_v = 24$ ($Q_{AVB}(95)$ ranged from 0.03 to 0.11). For all values of j_v the bias of VE estimates decreased as true VE increased (except for unbiased estimates when $j_v = 1$).

Study Length

We varied the length of the study period (15, 20, 25, 30 weeks) to determine if the window of the season captured by the study period influenced the bias of VE estimates. The season was only 57 weeks long therefore, we could not calculate the bias for 25 and 30 week study periods when vaccination started in week 38. For a fixed j_v, the length of the study period had a very small effect on the 95th quantile of absolute value of bias (Table 4.8). For a fixed study period length, we observed the same pattern of bias between the values of j_v as in the original analysis and other sensitivity bias. Specifically, bias was larger when vaccination started later in the outbreak, except under sources of bias A and C when the outcome of interest was SI.

Table 4.7: Bias of TN-based estimates of VE when true VE varies

Source of Bias[a]	**Outcome of Interest**	j_v	**True VE** 0.20	0.40	0.60	0.80
None	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.09	0.06	0.04	0.02
		38	0.16	0.12	0.08	0.04
A	SI & MAI	1	0.50	0.35	0.22	0.12
		24	0.38	0.28	0.19	0.09
		38	0.59	0.44	0.29	0.15
B1	SI & MAI	1	0.05	0.04	0.03	0.01
		24	0.09	0.06	0.04	0.02
		38	0.16	0.12	0.08	0.04
B2	SI & MAI	1	0.05	0.04	0.02	0.01
		24	0.12	0.09	0.06	0.02
		38	0.20	0.15	0.10	0.05
BS	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.09	0.06	0.04	0.02
		38	0.16	0.12	0.08	0.04
C[b]	SI	1	0.34	0.26	0.17	0.08
		24	0.11	0.09	0.06	0.03
		38	0.15	0.11	0.08	0.04
	MAI	1	0.00	0.00	0.00	0.00
		24	0.08	0.06	0.04	0.02
		38	0.15	0.12	0.08	0.04
D	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.09	0.06	0.04	0.02
		38	0.16	0.12	0.08	0.04

[a] For definitions of sources of bias, see Table 4.4. [b] Bias for estimates of VE against SI and MAI differ when bias C is present. The 95th quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, j_v, and true VE.

Table 4.8: Bias of TN-based estimates of VE when study length varies

Source of Bias[a]	Outcome of Interest	j_v	Study Length[b] 15	20	25	30
None	SI & MAI	1	0.00	0.00	0.00	0.00
		24	0.04	0.04	0.04	0.04
		38	0.05	0.07	-	-
A	SI & MAI	1	0.23	0.24	0.23	0.21
		24	0.19	0.17	0.16	0.16
		38	0.29	0.26	-	-
B1	SI & MAI	1	0.03	0.03	0.03	0.03
		24	0.04	0.04	0.04	0.04
		38	0.05	0.07	-	-
B2	SI & MAI	1	0.02	0.02	0.02	0.02
		24	0.05	0.05	0.05	0.06
		38	0.07	0.10	-	-
BS	SI & MAI	1	0.00	0.00	0.01	0.01
		24	0.04	0.04	0.04	0.05
		38	0.05	0.08	-	-
C[c]	SI	1	0.16	0.16	0.16	0.17
		24	0.07	0.05	0.05	0.04
		38	0.11	0.07	-	-
	MAI	1	0.00	0.00	0.00	0.01
		24	0.07	0.05	0.05	0.04
		38	0.05	0.07	-	-
D	SI & MAI	1	0.00	0.00	0.00	0.01
		24	0.04	0.04	0.04	0.05
		38	0.05	0.08	-	-

[a] For definitions of sources of bias, see Table 4.4. [b] We varied the length of the study period (15, 20, 25, 30 weeks). Bias for 25 and 30 week study periods for $j_v = 38$ could not be calculated because the season was only 57 weeks long. [c] Bias for estimates of VE against SI and MAI differ when bias C is present. The 95$^{\text{th}}$ quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, j_v, and study length.

4.2 Discussion

We assessed the bias of TN-based VE estimates when vaccination began at different weeks during an influenza outbreak by extending our model from Chapter 3. We used the 2009 pandemic as the motivating example, where a vaccine against the pandemic influenza strain was made available months after the beginning of the outbreak. Our results highlighted that VE estimates from TN studies suffer from bias when vaccination occurs during the study period, regardless of the source of bias present. However, the effect of the first week of vaccination was unpredictable. The largest increase in the magnitude of bias compared to the baseline scenario was 0.10. Median VE estimates for all values of j_v suffered from small to moderate under all sources of bias, except when vaccination affected the probability of NFARI (bias A) and vaccination reduced the probability of seeking medical care for FARI (bias C) and the outcome of interest was SI.

In some cases vaccination occuring during the study period mitigated the bias observed in the baseline scenario. For example, when vaccination affected the probability of NFARI (bias A), baseline VE estimates may suffer from severe bias ($Q_{AVB}(95) = 0.24$), but when vaccination occured during the study period, VE estimates suffered only substantial bias ($Q_{AVB}(95) = 0.13$ when $j_v = 24$). This phenomenon could be due to lower numbers of vaccinated controls when vaccination occurs during the study period relative to the baseline scenario. When all vaccinated individuals are vaccinated prior to the study period there would be a larger number of vaccinated controls biasing VE estimates downwards.

Overall, these results are incouraging. If the core assumption of the TN design is satisfied, that is, if bias A is not present, and VE estimates are not interpreted as VE against SI if it is suspected that vaccination reduces the probability of seeking care for FARI (due to reduced symptom severity), then TN-based VE estimates might only suffer from small to moderate bias.

4.3 All-or-None Vaccine Model

4.3.1 Background

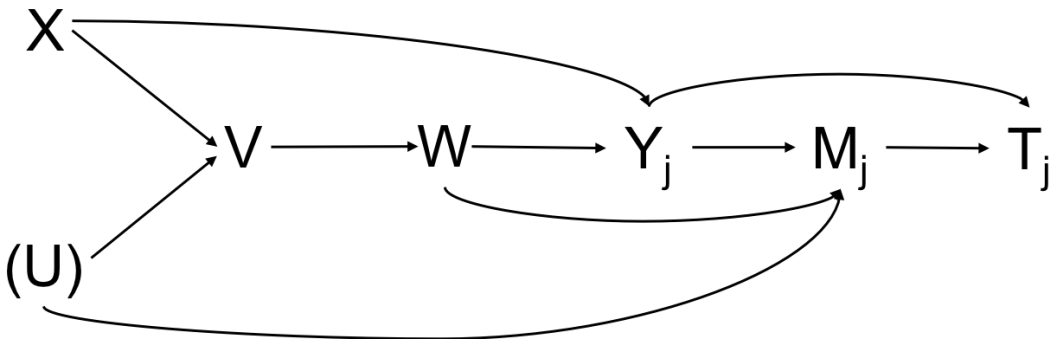
In Chapter 3, we assumed a leaky vaccine model in which vaccination provides a reduction in the probability of influenza infection [44]. An alternative vaccine model, called the all-or-none vaccine model can be applied to the influenza vaccine. In the all-or-none model, a proportion ρ of vaccinated individuals will acquire complete immunity to infection, while the remaining $1-\rho$ individuals will acquire no protection from vaccination [44]. In this way, the vaccine either provides full protection or no protection. To determine if the assumed vaccine model has an impact on the magnitude of bias of VE estimates from TN studies, we developed a second extension to the model presented in Chapter 3.

4.3.2 Methods

Model Description

We present an extension to the dynamic model presented in Chapter 3 in which we assume an all-or-none vaccine model (Figure 4.3). We introduce a new variable associated with protection from influenza vaccination. Details about the associated variables and the probabilities determining each variable's distribution can be found in Tables 4.9 and 4.10, respectively. We define a *standard person* as a person who is not protected by vaccination, healthy, and has high health awareness. A person is considered not protected by vaccination either by remaining unvaccinated or not acquiring protection after receipt of the vaccine.

Figure 4.3: Causal graph of influenza vaccine studies with covariates under all-or-none vaccine model.



X = health status, (U) = health awareness (unobserved), V = vaccination status, W = protection from vaccination, Y_j = ARI status in week j, M_j = seeking medical care for ARI in week j, and T_j = influenza test result in week j, where $j = 1, \ldots, J$ and J = the number of weeks in the study.

Step 1: Covariates. We assume that people within the population can be classified with a health status (X) of either "healthy" or "frail" and a health awareness (U) of either "high" or "low".

Step 2: Vaccination. Individuals are either vaccinated prior to the beginning of the study or remain unvaccinated. We use V to denote if a person received a vaccine (0 - unvaccinated, 1 - vaccinated). A proportion ρ of vaccinated individuals acquire complete protection from influenza infection, while the remaing $1-\rho$ vaccinated individuals acquire no protection. We let W indicate if a person acquires protection from vaccination, where $W = 0$ if a person remains unvaccinated or does not acquire protection and $W = 1$ if a person is vaccinated and acquires protection. We denote $\rho = \mathbb{P}(W = 1|V = 1)$. Whether or not a person acquires immunity from vaccination (W) is unobserved. We assume the distribution of W given V does not depend on a person's health status (X)

or health awareness (U).

Step 3: Influenza and non-influenza ARI. During the study, a person may become infected with an influenza virus and develop FARI and/or develop one or more NFARIs. We define a variable Y_j for the illness/infection status in week j, where the distribution of Y_j may depend on whether or not the person is protected by vaccination (W) and health status (X).

Step 4: Seeking medical care for ARI. A person with an ARI in week j may seek medical care (M_j). The probability of seeking medical care depends on Y_j, W, and U.

Step 5: Testing for influenza infection. We assume that each person who seeks medical care for ARI is tested for influenza infection (T_j).

Table 4.9: Variables in the model.

Variable	Definition	Values
X	Health status	0 - frail person 1 - healthy person
U	Health awareness (unobserved)	0 - low health awareness 1 - high health awareness
V	Vaccination status	0 - unvaccinated 1 - vaccinated
W	Protection from vaccination	0 - no 1 - yes
Y_j	Influenza/non-influenza ARI status in week j	0 - no ARI 1 - non-influenza ARI 2 - influenza ARI
M_j	Seeking medical care for ARI in week j	0 - no 1 - yes
T_j	Result of test for influenza infection in week j	0 - negative 1 - positive

Table 4.10: Parameters in the model.

Parameters	Definition	Comments
π_{xu}	$\mathbb{P}(X = x, U = u)$	
α_{xu}	$\mathbb{P}(V = 1 \mid X = x, U = u)$	
ρ	$\mathbb{P}(W = 1 \mid V = 1)$	
β_{jwx}	$\mathbb{P}(Y_j = 1 \mid W = w, X = x)$	$\beta_{j11} = \beta_{j01} \cdot \theta_\beta$
θ_β	multiplier for β when $W = 1$	$\beta_{j00} = \beta_{j01} \cdot \phi_\beta$
ϕ_β	multiplier for β when $X = 0$	$\beta_{j10} = \beta_{j01} \cdot \theta_\beta \cdot \phi_\beta$
γ_{jwx}	$\mathbb{P}(Y_j = 2 \mid W = w, X = x)$	$\gamma_{j11} = 0$
θ_γ	multiplier for γ when $W = 1$	$\gamma_{j00} = \gamma_{j01} \cdot \phi_\gamma$
ϕ_γ	multiplier for γ when $X = 0$	$\gamma_{j10} = 0$
δ_{1wu}	$\mathbb{P}(M_j = 1 \mid Y_j = 1, W = w, U = u)$	$\delta_{111} = \delta_{101} \cdot \theta_{\delta_1}$
θ_{δ_1}	multiplier for δ_1 when $W = 1$	$\delta_{100} = \delta_{101} \cdot \mu_{\delta_1}$
μ_{δ_1}	multiplier for δ_1 when $U = 0$	$\delta_{110} = \delta_{101} \cdot \theta_{\delta_1} \cdot \mu_{\delta_1}$
δ_{2wu}	$\mathbb{P}(M_j = 1 \mid Y_j = 2, W = w, U = u)$	$\delta_{211} = \delta_{201} \cdot \theta_{\delta_2}$
θ_{δ_2}	multiplier for δ_2 when $W = 1$	$\delta_{200} = \delta_{201} \cdot \mu_{\delta_2}$
μ_{δ_2}	multiplier for δ_2 when $U = 0$	$\delta_{210} = \delta_{201} \cdot \theta_{\delta_2} \cdot \mu_{\delta_2}$

A *standard person* is defined as a person with $(W = 0, X = 1, U = 1)$. β_{j01} and γ_{j01} represent the probabilities of contracting a NFARI or FARI for a *standard person* in week j, respectively, $j = 1, \ldots, J$. δ_{101} and δ_{201} represent the probabilities of seeking care for NFARI and FARI for a *standard person*, respectively. These probabilities, as well as all π_{xu}, $x = 0, 1; u = 0, 1$, all α_{jxu}, $j = 1, \ldots, J$; $x = 0, 1$; $u = 0, 1$, and all multipliers ($\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$), are input parameters. Under the all-or-none vaccine model, $\theta_\gamma = 0$ and $\theta_{\delta_2} = 0$.

True VE

True VE is defined as one minus the relative risk of influenza infection in vaccinated persons compared to unvaccinated persons. We define true VE against SI and MAI. Explicit expressions and derivations of true VE can be found Appendix C.2.1. A person is considered a true case of SI if s/he develops an ARI as a result of influenza infection ($Y_j = 2$). True VE against SI is defined as one minus the risk of influenza infection among vaccinated persons compared to

unvaccinated persons.

$$VE_{TSI} = 1 - \frac{\sum_{j=1}^{J} \mathbb{P}(Y_j = 2|V = 1)}{\sum_{j=1}^{J} \mathbb{P}(Y_j = 2|V = 0)} = \rho.$$

A person is considered a true case of MAI if s/he develops an influenza ARI during the study and seeks medical care for this ARI.

$$VE_{TMAI} = 1 - \frac{\sum_{j=1}^{J} \mathbb{P}(M_j = 1, Y_j = 2|V = 1)}{\sum_{j=1}^{J} \mathbb{P}(M_j = 1, Y_j = 2|V = 0)} = \rho.$$

Estimated VE

We assume that a person is classified as a TN case or a TN control at her/his first ARI-related visit. This classification does not change, regardless of possible conflicting test results in future visits. We also assume a person with no ARI does not seek care (i.e., if $Y_j = 0$, then $M_j = 0$).

A person is considered a case in week j, if they did not seek medical care for any ARI prior to week j ($\mathbf{M_{j-1}} = \mathbf{0}$), seek medical care for their ARI in week j ($M_j = 1$), and have FARI in week j ($Y_j = 2$). We can calculate the expected number of vaccinated and unvaccinated cases as

$$\mathbb{E}(\text{vaccinated cases in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, V = 1)$$

and

$$\mathbb{E}(\text{unvaccinated cases in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, V = 0)$$

A person is considered a control in week j in the same way as a case, except

that they have a NFARI in week j ($Y_j = 1$). The expected number of vaccinated and unvaccinated controls in week j can be calculated as

$$\mathbb{E}(\text{vaccinated controls in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, V = 1)$$

and

$$\mathbb{E}(\text{unvaccinated controls in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, V = 0)$$

Expressions of the expected numbers of cases and controls in terms of the model parameters are shown in Appendix C.2.2. To obtain the final 2x2 table for a TN study, we sum the expected cell counts over all weeks in the study period $1, \ldots, J$ (Table 4.11). The VE estimate from a TN study is

$$\hat{VE}_{TN} = 1 - \hat{OR}_{TN},$$

where $\hat{OR}_{TN}$ is the odds ratio from Table 4.11.

Table 4.11: Final 2x2 table for a TN study.

	Vaccinated	**Unvaccinated**
Case	$\sum_{j=1}^{J} \mathbb{E}(\text{vaccinated cases in week } j)$	$\sum_{j=1}^{J} \mathbb{E}(\text{unvaccinated cases in week } j)$
Control	$\sum_{j=1}^{J} \mathbb{E}(\text{vaccinated controls in week } j)$	$\sum_{j=1}^{J} \mathbb{E}(\text{unvaccinated controls in week } j)$

4.3.3 Calculations and Simulations

To evaluate the bias of VE estimates assuming an all-or-none vaccine model under different sources of bias (Table 4.4), we derived expressions of true and estimated VE from our model (Appendices C.2.1 and C.2.2). As in Chapter 3, we used these expressions of true and estimated VE to calculate the bias of VE

estimates under different sources of bias by varying the values of the corresponding parameters (Table 4.4). When a source of bias was absent, we kept the corresponding probability ratio fixed at 1.0. Bias was defined as estimated VE minus true VE. For bias A, we considered that protection from vaccination might increase or decrease the probability of NFARI, so θ_β varied from 0.5 to 2.0. For biases B1, B2, and BS, we allowed ϕ_β and ϕ_γ to vary between 0.5 and 1.0, since we expect healthy persons to have lower probabilities of ARI compared to frail persons. We did not assess bias C because the probability of a protected person seeking care for FARI is zero, $\theta_{\delta_2} = 0$. For bias D, we expect persons with high health awareness to have a higher probability of seeking medical care for both FARI compared to persons with low health awareness, so μ_{δ_2} varied between 0.5 to 1.0. We assume $\mu_{\delta_1} = \mu_{\delta_2}$.

We determined the 5$^{\text{th}}$, 50$^{\text{th}}$, and 95$^{\text{th}}$ quantiles of bias and 50$^{\text{th}}$ and 95$^{\text{th}}$ quantiles of absolute value of bias from 1,000 Monte Carlo simulations, as described in Section 4.1.2. Values of input parameters can be found in Appendix B.5.

We performed several sensitivity analyses to assess whether the patterns of bias observed in the main analysis were an artifact of our original input parameters. Specifically, we compared the bias of VE estimates under the all-or-none and leaky vaccine models for different values of true VE (0.2, 0.4, 0.6, and 0.8), vaccination coverage (0.2, 0.4, 0.6, and 0.8), when the probabilities of health status (X) and health awareness (U) were correlated, and for two alternative vaccination scenarios.

4.3.4 Results

We evaluated bias of TN-based VE estimates under the assumption of an all-or-none vaccine model in the presence of the sources of bias listed in Table 4.4.

Table 4.12 shows the 5^{th}, 50^{th}, and 95^{th} quantiles of bias and the 50^{th} and 95^{th} quantiles of absolute value of bias for each source of bias separately and then, for combinations of sources of bias. We compare the bias of estimates between the leaky vaccine model (assumed in Chapter 3) and the all-or-none vaccine model. When no sources of bias are present and under biases BS, D, and BS and D, TN-based estimates are unbiased under both vaccine models. Under source of bias A, the all-or-none estimate had a smaller 90% interval of bias (-0.11, 0.10) compared to the leaky estimate (-0.23, 0.15). Under bias B1 and B2, the magnitude of bias was similar between the two vaccine models; however, the leaky estimates had slightly smaller 90% intervals of bias compared to the all-or-none estimates.

4.3.5 Sensitivity Analyses

Alternative Values of True VE

To assess whether the value of true VE had an effect on the magnitude of bias under the all-or-none vaccine model, we varied the value of true VE (0.2, 0.4, 0.6, and 0.8). Table 4.13 compares the 95^{th} quantile of absolute value of bias of VE estimates under the leaky and all-or-none vaccine models in the presence of different sources of bias when the true VE varies. As true VE increases, the difference in the magnitude of bias between the leaky and all-or-none estimates decreases, becoming small (<0.05) when true VE$\geq$0.60. Additionally, for both vaccine models, as the true VE increases, the magnitude of bias decreases in the presence of biases B1 and B2. Interestingly, this decreasing trend is not observed for the all-or-none model in the presence of bias A. Finally, TN-based VE

Table 4.12: Comparison of bias of TN-based estimates between leaky and all-or-none vaccine models

Source of Bias[a]	Vaccine Model	$Q_B(5)$[b]	$Q_B(50)$	$Q_B(95)$	$Q_{AVB}(50)$	$Q_{AVB}(95)$
None	Leaky	0.00	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00	0.00
A	Leaky	-0.23	0.00	0.15	0.08	0.23
	All-or-None	-0.11	0.00	0.10	0.05	0.12
B1	Leaky	0.00	0.02	0.04	0.02	0.04
	All-or-None	0.01	0.03	0.07	0.03	0.07
B2	Leaky	-0.04	-0.01	0.00	0.01	0.04
	All-or-None	-0.07	-0.02	0.00	0.02	0.07
BS	Leaky	0.00	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00	0.00
D	Leaky	0.00	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00	0.00
BS, D	Leaky	0.00	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00	0.00
BS, D, A	Leaky	-0.25	0.00	0.15	0.09	0.25
	All-or-None	-0.11	0.01	0.10	0.05	0.12

[a] Sources of bias: A - vaccination affects the probability of NFARI, B1 - Healthy persons have a lower probability of NFARI, B2 - Healthy persons have a lower probability of FARI, BS - Healthy persons have a lower probability of NFARI and FARI, and D - ARI patients with high health awareness have a higher probability of seeking medical care.

[b] $Q_B(5)$=5th quantile, $Q_B(50)$=50th quantile (median), $Q_B(95)$=95th quantile, $Q_{AVB}(50)$=50th quantile of the absolute value of bias, $Q_{AVB}(95)$=95th quantile of the absolute value of bias. Quantiles were determined from 1,000 Monte Carlo simulations. Absolute value of bias is defined as the difference between the estimate VE and the true VE without regard to the sign.

Table 4.13: Bias of TN-based estimates of VE when true VE varies

Source of Bias	Vaccine Model	True VE 0.20	0.40	0.60	0.80
None	Leaky	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00
A	Leaky	0.30	0.23	0.15	0.08
	All-or-None	0.08	0.12	0.12	0.08
B1	Leaky	0.07	0.05	0.04	0.02
	All-or-None	0.13	0.10	0.07	0.03
B2	Leaky	0.07	0.06	0.04	0.02
	All-or-None	0.15	0.11	0.07	0.04
BS	Leaky	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00
D	Leaky	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00

For definitions of sources of bias, see Table 4.4. The 95th quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, vaccine model, and true VE.

estimates are unbiased under no source of bias and biases BS and D regardless of the vaccine model.

Vaccination Coverage

To assess whether the differences in magnitude of bias between the leaky and all-or-none vaccine models was attributable to the vaccination coverage, we varied the value of vaccination coverage (Table 4.14). Vaccination coverage did not affect the magnitude of bias of all-or-none VE estimates. However, under biases B1 and B2 the magnitude of bias of leaky VE estimates increased as vaccination coverage increased. When vaccination coverage was 80% leaky VE estimates suffered from a larger bias than all-or-none estimates. In the presence of bias A, leaky estimates suffered from severe bias, whereas all-or-none VE estimates suffered from substantial bias. Under no source of bias and biases BS

Table 4.14: Bias of TN-based estimates of VE when vaccination coverage varies

Source of Bias	Vaccine Model	Vaccination Coverage 0.20	0.40	0.60	0.80
None	Leaky	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00
A	Leaky	0.34	0.30	0.35	0.33
	All-or-None	0.12	0.11	0.12	0.12
B1	Leaky	0.03	0.04	0.06	0.12
	All-or-None	0.07	0.07	0.07	0.07
B2	Leaky	0.03	0.04	0.07	0.15
	All-or-None	0.07	0.07	0.07	0.08
BS	Leaky	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00
D	Leaky	0.00	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00	0.00

For definitions of sources of bias, see Table 4.4. The 95th quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, vaccine model, and vaccination coverage.

and D, VE estimates from both vaccine models were unbiased for regardless of vaccination coverage.

Alternative Probabilities of (X,U)

Table 4.15 compares the 95th quantile of absolute value of bias of VE estimates under the leaky and all-or-none vaccine models when the probabilities of (X, U) are independent and correlated. The same pattern of bias between the leaky and all-or-none vaccine models was observed when the probabilities of (X, U) were correlated compared to when the probabilities of (X, U) were independent. Bias of VE estimates under the all-or-none model were very similar regardless of the probilities of (X, U). Under the leaky model, bias of VE estimates varied in the presence of biases B1 and B2. Compared to when the probabilities of health status and health awareness were independnent, smaller bias was

Table 4.15: Bias of TN-based estimates of VE for alternative probabilities of (X,U)

Source of Bias	Vaccine Model	Probabilities of (X,U)		
		Independent	Alternative 1[a]	Alternative 2[b]
None	Leaky	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00
A	Leaky	0.34	0.32	0.33
	All-or-None	0.12	0.12	0.12
B1	Leaky	0.05	0.03	0.09
	All-or-None	0.07	0.06	0.08
B2	Leaky	0.05	0.02	0.10
	All-or-None	0.07	0.07	0.09
BS	Leaky	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00
D	Leaky	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00

[a] The probability of high health awareness given a healthy health status is greater than the probability of having high health awareness given frail health status ($\mathbb{P}(U = 1|X = 1) > \mathbb{P}(U = 1|X = 0)$).
[b] The probability of high health awareness given a healthy health status is less than the probability of having high health awareness given frail health status ($\mathbb{P}(U = 1|X = 1) < \mathbb{P}(U = 1|X = 0)$).The 95$^{\text{th}}$ quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, vaccine model, and probability of (X, U).

observed when the probability of high health awareness given a healthy health status is greater than the probability of having high health awareness given frail health status (alternative 1), and larger bias was observed when the probability of high health awareness given a healthy health status is less than the probability of having high health awareness given frail health status (alternative 2).

Table 4.16: Bias of TN-based estimates of VE for alternative probabilities of vaccination

Source of Bias	Vaccine Model	Probabilities of Vaccination		
		Independent	Alternative 1[a]	Alternative 2[b]
None	Leaky	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00
A	Leaky	0.34	0.35	0.33
	All-or-None	0.12	0.12	0.12
B1	Leaky	0.05	0.05	0.02
	All-or-None	0.07	0.03	0.06
B2	Leaky	0.05	0.05	0.02
	All-or-None	0.07	0.03	0.06
BS	Leaky	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00
D	Leaky	0.00	0.00	0.00
	All-or-None	0.00	0.00	0.00

[a] The probability of being vaccinated was lower in persons with frail health status ($X = 0$), low health awareness ($U = 0$), or both ($X = 0, U = 0$). [b] The probability of being vaccinated much higher in persons with frail health status and high health awareness ($X = 0, U = 1$). The 95$^{\text{th}}$ quantile of absolute value of bias from 1,000 Monte Carlo simulations is reported for each combination of source of bias, vaccine model, and probability of vaccination.

Alternative Probabilities of Vaccination

Table 4.16 compares the 95$^{\text{th}}$ quantile of absolute value of bias of VE estimates under the leaky and all-or-none vaccine models under different probablities of vaccination. The same pattern of bias between the leaky and all-or-none vaccine models was observed when the probabilities of vaccination were and were not multiplicative. Additionally, within a given vaccine model the bias did not differ greatly between the three vaccination scenarios.

4.3.6 Discussion

In this section we evaluated the bias of influenza VE estimates from TN studies under the assumption of an all-or-none vaccine model, a special case of the vaccine model assumed in Chapter 3. The primary findings of this work are: (1) under the assumption of an all-or-none model, TN-based estimates of VE are less susceptible to bias when vaccination affects the probability of NFARI (bias A), (2) when health status impacts the probability of NFARI (bias B1) and FARI (bias B2) VE estimates under the all-or-none vaccine model suffered only small increases in bias compared to VE estimates under the leaky vaccine model, and (3) TN-based estimates were unbiased when health status had the same affect on the probabilities of NFARI and FARI (bias BS) and health awareness affected the probability of seeking care for ARI (bias D) regardless of the assumed vaccine model.

The phenomenon from finding (1) is likely due to our assumption that the probability of developing NFARI depends on whether or not a person was protected by vaccination (W) rather than on receipt of the vaccine (V). Under the all-or-none vaccine model only a proportion of individuals who receive the vaccine are protected thus, fewer individuals are impacted by bias A compared to the leaky vaccine model, resulting in less drastic changes in VE. However, despite the smaller magnitude of bias, VE estimates may still suffer from substantial bias when vaccination affects the probability of NFARI. The vaccine model is usually unknown, so despite the reduction in bias under the all-or-none model, the use of the TN design is not recommended if there is evidence to support that vaccination affects the probability of NFARI.

We performed several sensitivity analyses and observed the same pattern

of bias between the leaky and all-or-none model within each sensitivity analysis. However, the difference between the bias for leaky and all-or-none VE estimates diminished as true VE increased, becoming small when true VE was 80%. From this, we conclude that for highly effective vaccines the assumption of vaccine model has a very small impact on the magnitude of bias.

In this work, we assumed that the vaccine model was the same for every individual. It is conceivable that a hybrid vaccine model may be more realistic, in which the vaccine model is not the same for every individual and instead results in one of three outcomes: complete protection, a reduction in the probability of influenza infection (incomplete protection), and no protection. An individual's response to the vaccine likely depends on covariates, such as health status, prior vaccination, or prior infection. In this work, we assumed that the probability of complete protection from vaccination given receipt of the vaccine did not depend on any covariates. An extension of this model could allow the probability of protection to depend on covariates, allowing the assessment of the bias of TN-based estimates under a hybrid vaccine model. Additionally, this work only focuses on assessing the bias of VE estimates from a TN study because of the popularity of this study design for assessing influenza VE. Future work should investigate the impact that vaccine model has on VE estimates from other study designs. Finally, this work only considers unadjusted estimates of VE. Assessment of common methods (e.g., logistic regression) for adjusting VE estimates to mitigate bias should be evaluated to establish how effective these methods are at controlling for sources of bias and confounding.

Chapter 5

Discussion

5.1 Work In Progress: Alternative case and control definitions for test-negative studies

In Chapter 3 we presented a model to quantify the bias of VE estimates from four different types of observational studies. In Chapter 4, we extended this model to further evaluate the bias of VE estimates from TN studies when vaccinations occurred during the study period and under an all-or-none vaccine model. However, in the models presented in the previous chapters we consider only a person's first ARI-related medical visit to determine their case or control status for the study designs requiring outpatient medical care as an inclusion criterion. It is possible for a person to seek medical care for ARI multiple times throughout the duration of the study period, particularly during the months when influenza viruses and other non-influenza viruses are in peak circulation. If a person makes numerous ARI-related medical visits during a study, their case or control status may be defined differently. To determine the effect of different definitions of cases and controls within TN studies, we have developed a third extension of the model presented in Chapter 3.

5.1.1 Background

Cases and controls in TN studies are determined by the results of being tested for influenza infection after seeking medical care. In Chapter 3, cases and controls were determined at a person's first medical visit for ARI, and any subsequent visits were ignored. However, throughout the study period, a person may seek care for ARI more than once. For example, consider an individual who seeks care for ARI twice and tests negative for influenza during their first visit and tests positive during their second visit. This person could be included in the study in several ways: only as a control (because they tested negative first), only as a case (because they had a positive test result), or as a control for their first visit and as a case for their second visit (i.e., they are counted separately for each visit).

In a 2013 study, De Serres *et al.* investigated TN-based estimates of VE from RCT data using different definitions of TN controls. Specifically, the study compared estimates of VE when controls were defined as (a) participants with any negative swabs (regardless of any positive swabs at another time); (b) participants with only negative swabs (who never tested positive); and (c) all NFARI episodes (a participant could be counted more than once). The study found little difference in TN-based estimates of VE regardless of the definition of controls used [45]. However, the study did not assess TN-based estimates with respect to bias or consider alternative definitions of cases. Furthermore, the use of RCT data does not account for possible confounding present in an actual TN study. We present a model that allows for more than one ARI-related medical visit during a TN study period. Using this model, we aim to assess how different definitions of cases and controls influence the bias of TN-based VE estimates.

Methods

We use the same model presented in Chapter 3, where each person in our study population has a health status (X) and a health awareness (U), and may be vaccinated prior to the beginning of the study (V), develop an ARI in week j (Y_j), and seek care for that ARI in week j (M_j). Unlike in Chapter 3, where we ignored any subsequent medical visits after the first vist, we assume a person seeks medical care for ARI at most twice during the study period. At each visit they are tested for influenza infection (T_j). We assume the influenza test has perfect sensitivity and specificity.

We first calculate the probability of a given outcome or combination of outcomes in each week and then use these probabilities to determine the expected number of cases and controls. Using the expected numbers of cases and controls we can obtain an estimate of VE as one minus the odds ratio.

Since the event of seeking care once during the study is mutually exclusive from seeking care twice, we can first determine the probability of seeking care once for either FARI or NFARI. We assume a person with no ARI does not seek care. The probability of seeking care once during the study is the same as the probability of developing an ARI and seeking care in week j ($Y_j = y_j, M_j = 1$) and not seeking care in any other week ($\mathbf{M_{J^*}} = \mathbf{0}$), where J^* is the set $\{1, \ldots, J\}$, excluding j. We can express the probability of one ARI-related medical visit for one j and combination of (v, x, u) as

$$
\begin{aligned}
&\mathbb{P}(M_j = 1, Y_j = y_j, \mathbf{M_{J^*}} = \mathbf{0} | V = v, X = x, U = u) \\
&= \sum_{(Y_1, \ldots, Y_J)} (\mathbb{P}(M_j = 1, Y_j = y_j, \mathbf{M_{J^*}} = \mathbf{0} | Y_1, \ldots, Y_J, V = v, X = x, U = u) \\
&\times \qquad \mathbb{P}(Y_1, \ldots, Y_J | V = v, X = x, U = u)) \,,
\end{aligned}
$$

where $y_j = 1, 2$. Since $M_1, \ldots, M_J$ are independent given $Y_1, \ldots, Y_J$,

$$\begin{aligned}
&\mathbb{P}(M_j = 1, Y_j = y_j, \mathbf{M_{J^*}} = \mathbf{0} | Y_1, \ldots, Y_J, V = v, X = x, U = u) \\
&= \mathbb{P}(M_j = 1 | Y_j = y_j, V = v, X = x, U = u) \prod_{k \in J^*} \mathbb{P}(M_k = 0 | Y_k = y_k, V = v, X = x, U = u).
\end{aligned}$$

To determine the joint probability of ARI status in each week $(Y_1, \ldots, Y_J | V = v, X = x, U = u)$, consider week j_F, $j_F = 1, \ldots, J+1$ to be the week in which a person has FARI ($Y_{j_F} = 2$). The joint probility of ARI status for every week can be written as

$$\begin{aligned}
&\mathbb{P}(Y_1, \ldots, Y_J | V = v, X = x, U = u) \\
&= \mathbb{P}(Y_1, \ldots, Y_J | Y_{j_F} = 2, V = v, X = x, U = u) \mathbb{P}(Y_{j_F} = 2 | V = v, X = x, U = u)
\end{aligned}$$

By conditioning on $Y_{j_F} = 2$, we can treat the ARI status in the remaining weeks as independent, since the probability of NFARI is independent from one week to the next. Thus,

$$\begin{aligned}
&\mathbb{P}(Y_1, \ldots, Y_J | Y_{j_F} = 2, V = v, X = x, U = u) \\
&= \prod_{h=1}^{J} \mathbb{P}(Y_h < 2 | Y_{j_F} = 2, V = v, X = x, U = u)
\end{aligned}$$

For each value of j the joint probability of seeking care once for ARI and having vaccination status v is

$$\begin{aligned}
&\mathbb{P}(M_j = 1, Y_j = y_j, \mathbf{M_{J^*}} = \mathbf{0}, V = v) \\
&= \sum_{x=0}^{1} \sum_{u=0}^{1} [\mathbb{P}(M_j = 1, Y_j = y_j, \mathbf{M_{J^*}} = \mathbf{0} | V = v, X = x, U = u) \\
&\times \qquad \mathbb{P}(V = v | X = x, U = u) \mathbb{P}(X = x, U = u)]
\end{aligned}$$

Finally, the overall probability of seeking care once during the study period for ARI status y_j and vaccination status v is the sum over all weeks $j = 1, \ldots, J$.

To determine the probability of seeking care twice during the study period, we developed an algorithm to determine all possible combinations of ARI outcomes given a person seeks care twice. This algorithm can be extended to accomodate more than two medical visits; however, we leave this to future work.

We assume a person makes two visits to seek medical care in weeks j_1 and j_2 and has ARI status $Y_{j_1} = y_{j_1}$ and $Y_{j_2} = y_{j_2}$, respectively. Since we assume a person can have at most one FARI during the study, there are three possible outcomes for which a person seeks medical twice: having NFARI in week j_1 and FARI in week j_2 ($Y_{j_1} = 1, Y_{j_2} = 2$), having FARI in week j_1 and NFARI in week j_2 ($Y_{j_1} = 2, Y_{j_2} = 1$), or having NFARI twice in weeks j_1 and j_2 ($Y_{j_1} = 1, Y_{j_2} = 1$). Using the following algorithm we can determine the probability of seeking medical care for ARI twice during the study period for a given (v, x, u).

1. Specify outcome:

   $(Y_{j_1} = 1, Y_{j_2} = 2 | V = v, X = x, U = u)$,

   $(Y_{j_1} = 2, Y_{j_2} = 1 | V = v, X = x, U = u)$, or

   $(Y_{j_1} = 1, Y_{j_2} = 1 | V = v, X = x, U = u)$

2. Specify weeks of visits: $(j1, j2)$, where $j1 < j2$

3. Specify $(Y_1, \ldots, Y_J | V = v, X = x, U = u)$

4. Calculate $\mathbb{P}(M_{j_1} = 1, Y_{j_1} = y_{j_1}, M_{j_2} = 1, Y_{j_2} = y_{j_2}, \mathbf{M_{J^{**}}} = \mathbf{0} | Y_1, \ldots, Y_J, V = v, X = x, U = u) \times \mathbb{P}(Y_1, \ldots, Y_J | V = v, X = x, U = u)$, where $y_{j_1} = 1, 2$, $y_{j_2} = 1, 2$, and $J^{**} = \{1, \ldots, J\}$ excluding j_1, j_2.

5. Repeat steps (3) and (4) for each combination of
   $(Y_1, \ldots, Y_J | V = v, X = x, U = u)$
6. Sum probabilities from (4) for every combination of
   $(Y_1, \ldots, Y_J | V = v, X = x, U = u)$
7. Repeat steps (2) thru (6) for each pair (j_1, j_2), $j_1 < j_2$
8. Sum probabilities from (6) over all possible pairs (j_1, j_2)

The joint probability of seeking medical in weeks j_1 and j_2 can be written as the probability of seeking medical care in weeks j_1 and j_2 and not seeking care in any of the remaining weeks. Thus, for a single pair (j_1, j_2) and combination of (v, x, u),

$$\begin{aligned}
&\mathbb{P}(M_{j_1} = 1, Y_{j_1} = y_{j_1}, M_{j_2} = 1, Y_{j_2} = y_{j_2}, \mathbf{M_{J^{**}}} = \mathbf{0} | V = v, X = x, U = u) \\
&= \sum_{(Y_1, \ldots, Y_J)} [\mathbb{P}(M_{j_1} = 1, Y_{j_1} = y_{j_1}, M_{j_2} = 1, Y_{j_2} = y_{j_2}, \mathbf{M_{J^{**}}} = \mathbf{0} | Y_1, \ldots, Y_J, V = v, X = x, U = u) \\
&\times \qquad \mathbb{P}(Y_1, \ldots, Y_J | V = v, X = x, U = u)] ,
\end{aligned}$$

where $y_{j_1} = 1, 2$, $y_{j_2} = 1, 2$, and $J^{**} = \{1, \ldots, J\}$ excluding j_1, j_2. Since $M_1, \ldots, M_J$ are independent given $Y_1, \ldots, Y_J$,

$$\begin{aligned}
&\mathbb{P}(M_{j_1} = 1, Y_{j_1} = y_{j_1}, M_{j_2} = 1, Y_{j_2} = y_{j_2}, \mathbf{M_{J^{**}}} = \mathbf{0} | Y_1, \ldots, Y_J, V = v, X = x, U = u) \\
&= \mathbb{P}(M_{j_1} = 1 | Y_{j_1} = y_{j_1}, V = v, X = x, U = u) \mathbb{P}(M_{j_2} = 1 | Y_{j_2} = y_{j_2}, V = v, X = x, U = u) \\
&\times \prod_{k \in J^{**}} \mathbb{P}(M_k = 0 | Y_k = y_k, V = v, X = x, U = u).
\end{aligned}$$

For each pair (j_1, j_2) the joint probability of seeking care twice for ARI and having vaccination status v is

$$\begin{aligned}
&\mathbb{P}(M_{j_1} = 1, Y_{j_1} = y_{j_1}, M_{j_2} = 1, Y_{j_2} = y_{j_2}, \mathbf{M_{J^{**}}} = \mathbf{0}, V = v) \\
&= \sum_{x=0}^{1} \sum_{u=0}^{1} [\mathbb{P}(M_{j_1} = 1, Y_{j_1} = y_{j_1}, M_{j_2} = 1, Y_{j_2} = y_{j_2}, \mathbf{M_{J^{**}}} = \mathbf{0} | V = v, X = x, U = u) \\
&\times \qquad \mathbb{P}(V = v | X = x, U = u) \mathbb{P}(X = x, U = u)]
\end{aligned}$$

Finally, the overall probability of seeking care twice during the study period for outcome (y_{j_1}, y_{j_2}) and vaccination status v is the sum over all possible pairs of (j_1, j_2).

We intend to compare TN-based estimates of VE, using these probabilities, under five different definitions of cases and controls:

1. Case/control status is based on a person's first medical visit and all subsequent visits are ignored (original definition used in Chapter 3).

2. A case is a person who tested positive for influenza at least once, and a control is a person who had at least one negative test result and no positive test results.

3. A case is a person who tested positive for influenza at least once and a person is classified as a control if (a) s/he had only negative test results, or (b) s/he had at least one negative test result before any positive test results.

   A person who had one or more negative results followed by a positive result is counted both as a case and a control; however, negative results after the first positive result are ignored (censored).

4. A case is a person who tested positive for influenza at least once and a control is a person who tested negative at least once, regardless of any prior or subsequent positive test results (uncensored).

5. Each visit contributes one observation. A case is a visit resulting in a positive test result and a control is a visit resulting in a negative test result. A person may be included more than once in the study.

The expected number of vaccinated and unvaccinated cases and controls can be calculated as N times the corresponding probabilities, where N is the population size. For example, for definition (2), the expected number of vaccinated cases and controls are

$$\begin{aligned}
&\mathbb{E}(\text{vaccinated cases}) \\
&= N \times \left(\sum_{j=1}^{J} \mathbb{P}(M_j = 1, Y_j = 2, \mathbf{M_{J^*}} = \mathbf{0}, V = 1) \right. \\
&+ \sum_{(j_1, j_2)} \mathbb{P}(M_{j_1} = 1, Y_{j_1} = 1, M_{j_2} = 1, Y_{j_2} = 2, \mathbf{M_{J^{**}}} = \mathbf{0}, V = 1) \\
&+ \left. \sum_{(j_1, j_2)} \mathbb{P}(M_{j_1} = 1, Y_{j_1} = 2, M_{j_2} = 1, Y_{j_2} = 1, \mathbf{M_{J^{**}}} = \mathbf{0}, V = 1) \right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}(\text{vaccinated controls}) \\
&= N \times \left(\sum_{j=1}^{J} \mathbb{P}(M_j = 1, Y_j = 1, \mathbf{M_{J^*}} = \mathbf{0}, V = 1) \right. \\
&+ \left. \sum_{(j_1, j_2)} \mathbb{P}(M_{j_1} = 1, Y_{j_1} = 1, M_{j_2} = 1, Y_{j_2} = 1, \mathbf{M_{J^{**}}} = \mathbf{0}, V = 1) \right)
\end{aligned}$$

We will compare the bias of VE estimates from TN studies under the sources of bias described in Chapter 3 (Section 3.2.4) to determine if the TN design is robust to different definition of cases and controls. We will validate our model

using stochastic simulations. We hope this work can better inform the design and implementation of TN studies, producing more reliable VE estimates.

5.2 Future Work

In the preceding chapters, the main focus has been on unadjusted VE estimates first, the estimation of unadjusted source-specific estimates of influenza VE (Chapter 2) and second, the evaluation of bias of unadjusted VE estimates from different observational studies in the presence of different sources of bias (Chapters 3 and 4). With regard to bias, our main objective was to characterize the bias of VE estimates from different sources rather than to identify or develop methods to adjust for this bias. This work highlights that bias is present in observational studies for the estimation of influenza VE. Depending on the study design, this bias may be a result of differing health-care seeking behavior between vaccinees and non-vaccinees, the effect of vaccination on the probability of developing NFARI, the effect of health status on both the probability of vaccination and the probability of FARI, or misclassification of influenza infection status. While some bias can be mitigated through careful study design (for example employing an ASC study if vaccination affects the probability of NFARI), other sources of bias must be adjusted for after the study has been conducted. Thus, it is important to identify statistical techniques to control for the bias of VE estimates. Previous work has identified sources of bias [122, 123] and characterized the magnitude of this bias [99, 124], but little work has been done to quantitatively assess and compare methods for bias adjustment. Future work should focus on developing methods and techniques to successfully reduce the bias in VE estimates identified in Chapters 3 and 4.

An important confounder not usually accounted for in VE studies, and not considered in this work, is repeated influenza vaccination or prior influenza infection [122]. Repeated influenza vaccination has received growing interest as more and more people are annually immunized [126]. Historically, studies have reported inconsistent findings regarding whether repeated vaccination is detrimental to protection against influenza infection [127–129]. Several recent studies have found lower VE among individuals who were vaccinated in the current season and the prior season compared to individuals only vaccinated in the current season [94, 130, 131]. A 2017 systematic review and meta-analysis found no evidence that prior season vaccination negatively impacts current season VE [126]. Despite these studies, there has been no theoretical assessment of the potential bias of VE estimates caused by repeated vaccination or past influenza infection. Like prior vaccination, past influenza infection may effect the probability of influenza infection in subsequent seasons [132, 133]. The models presented in this dissertation could be further extended to incorporate vaccination or influenza infection status in prior seasons to better characterize how vaccination and infection history impacts current VE estimates.

Another important consideration in the evaluation of influenza VE is waning immunity from vaccination over time. Several studies conducted in Europe during the 2011-2012 influenza season found that vaccinated individuals tended to present with influenza infection later in the season, suggesting some degree of waning immunity from vaccination [134–137]. Since the 2011-2012 season, studies conducted in Australia [138], the United Kingdom [139], Europe [140], and the Unites States [141] have found conflicting results regarding intraseason waning. A recent study pooled data from the 2011-2012 through

2014-2015 influenza seasons and observed decreasing influenza vaccine protection as time since vaccination increased [141]. However, this study only showed an association between VE and time since vaccination, not causation. If intraseason waning of protection from the influenza vaccine is a true biological phenomenon, it is important to consider the impact of waning immunity on estimates of VE.

Finally, we compare different types of observational studies with respect to bias and make recommendations about which study design is most appropriate in a given set of circumstances. However, we do not account for the cost associated with these studies. In practice, cost is a major consideration in the choice of study design. Future work should include a cost-benefit analysis of different types of observational studies to determine if the benefit of using a more expensive study design, such as an ASC study, is justified by the improvement in estimates of VE.

Appendix A

Appendix to Chapter 2

A.1 Source-Specific Probabilities of Infection

To estimate source-specific VE, we define λ_{iH} and λ_{iC} as the probability that person i is infected from a household (H) or community (C) contact during the study, respectively. The source-specific probabilities of infection during the study can be written in terms of the daily probabilities of infection. The source-specific probabilities of being infected on day d are:

$$\mathbb{P}(\text{Infected from the household on day } d) = [1-(1-\beta_v)^{m_{id}}]\,(1-\gamma_v p(d))\psi_{i0(d-1)}$$
$$\mathbb{P}(\text{Infected from the community on day } d) = (1-\beta_v)^{m_{id}}\gamma_v p(d)\psi_{i0(d-1)}$$

The probabilities of being infected during the study can be expressed as the sum of the daily probabilities of infection,

$$\begin{aligned}\lambda_{iH} &= \sum_{d=1}^{D}\left\{[1-(1-\beta_v)^{m_{id}}]\,(1-\gamma_v\cdot p(d))\cdot\psi_{i0(d-1)}\right\}\\ \lambda_{iC} &= \sum_{d=1}^{D}\left\{(1-\beta_v)^{m_{id}}\gamma_v\cdot p(d)\cdot\psi_{i0(d-1)}\right\}\end{aligned}$$

When estimating the VEs we ignore the possibility of co-infection and assume that a person can only be infected once during the study.

A.2 Model Adequacy

Table A.1: Frequencies of cases per household size from 200 simulated influenza outbreaks using the Michigan Study data as input

# of cases / **HH Size**	0	1	2	3	4	5	6
4	124.70	30.87	8.36	1.83	0.25	0	0
5	71.07	19.86	6.09	1.57	0.38	0.04	0
6	25.60	7.73	2.61	0.83	0.21	0.04	0
7	8.63	2.59	1.09	0.48	0.19	0.05	0.01
8	2.45	0.96	0.38	0.15	0.05	0.02	0
9	0	0	0	0	0	0	0
10	1.04	0.37	0.25	0.14	0.12	0.07	0.02

Using the estimated transmission parameters from the data analysis as input for our simulation program, we obtained mean number of cases per household size from 200 influenza outbreak simulations.

Table A.2: Observed frequencies of cases per household size from the Michigan Study

# of cases / **HH Size**	0	1	2	3	4	5	6
4	129	26	8	2	1	0	0
5	77	17	3	2	0	0	0
6	25	8	1	1	2	0	0
7	9	3	0	0	0	1	0
8	3	1	0	0	0	0	0
9	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0

Appendix B

Appendix to Chapter 3

B.1 True VE

B.1.1 True VE Against SI

A person is considered a true case of SI if s/he develops an ARI as a result of influenza infection. For a given vaccination status v the probability of SI in week j can be written as follows:

$$\begin{aligned}\mathbb{P}(Y_j = 2|V = v) &= \sum_{x=0}^{1} \mathbb{P}(Y_j = 2|V = v, X = x) \times \mathbb{P}(X = x|V = v) \\ &= \sum_{x=0}^{1} \gamma_{jvx} \times \mathbb{P}(X = x|V = v)\end{aligned}$$

Because of assumption (d) above, the probability of SI over the entire study given vaccination status v is

$$\mathbb{P}(\text{person classified as case of SI}|V = v) = \sum_{j=1}^{J} \sum_{x=0}^{1} \left[\gamma_{jvx} \times \mathbb{P}(X = x|V = v)\right].$$

where

$$\begin{aligned}\mathbb{P}(X=x|V=v) &= \frac{\mathbb{P}(V=v|X=x)\mathbb{P}(X=x)}{\mathbb{P}(V=v)} \\ &= \frac{\sum_{u=0}^{1}\mathbb{P}(V=v|X=x,U=u)\mathbb{P}(X=x,U=u)}{\sum_{x=0}^{1}\sum_{u=0}^{1}\mathbb{P}(V=v|X=x,U=u)\mathbb{P}(X=x,U=u)}.\end{aligned}$$

Using our model parameters we can rewrite $\mathbb{P}(X=x|V=v)$ as

$$\begin{aligned}\mathbb{P}(X=0|V=0) &= \frac{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\ \mathbb{P}(X=1|V=0) &= \frac{(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\ \mathbb{P}(X=0|V=1) &= \frac{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\ \mathbb{P}(X=1|V=1) &= \frac{\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}}\end{aligned}$$

When evaluating true VE we assume random vaccination, i.e. vaccination does not depend on health status X or health awareness U ($\alpha_{xu}=\alpha\ \forall x,u$). Thus, the above expressions of $\mathbb{P}(X=x|V=v)$ can be simplified as

$$\begin{aligned}\mathbb{P}(X=0|V=0) &= \frac{(1-\alpha)(\pi_{00}+\pi_{01})}{(1-\alpha)(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{00}+\pi_{01} \\ \mathbb{P}(X=1|V=0) &= \frac{(1-\alpha)(\pi_{10}+\pi_{11})}{(1-\alpha)(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{10}+\pi_{11} \\ \mathbb{P}(X=0|V=1) &= \frac{\alpha(\pi_{00}+\pi_{01})}{\alpha(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{00}+\pi_{01} \\ \mathbb{P}(X=1|V=1) &= \frac{\alpha(\pi_{10}+\pi_{11})}{\alpha(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{10}+\pi_{11}\end{aligned}$$

Therefore, the true VE against SI (VE_{TSI}) can be written as one minus the ratio of the probabilities of SI in vaccinated and unvaccinated persons:

$$VE_{TSI} = 1 - \frac{\sum_{j=1}^{J} [\gamma_{j10}(\pi_{00} + \pi_{01}) + \gamma_{j11}(\pi_{10} + \pi_{11})]}{\sum_{j=1}^{J} [\gamma_{j00}(\pi_{00} + \pi_{01}) + \gamma_{j01}(\pi_{10} + \pi_{11})]}.$$

B.1.2 True VE Against MAI

A person is considered a true case of MAI if s/he develops an ARI as a result of influenza infection and seeks medical care for their ARI. Similarly to SI, the probability of MAI given vaccination v for a given week j can be written as:

$$\begin{aligned}
&\mathbb{P}(\text{person considered case of MAI in week } j | V = v) \\
&= \sum_{x=0}^{1} \sum_{u=0}^{1} \mathbb{P}(Y_j = 2, M_j = 1 | V = v, X = x, U = u) \times \mathbb{P}(X = x, U = u | V = v) \\
&= \sum_{x=0}^{1} \sum_{u=0}^{1} \delta_{2vu} \gamma_{jvx} \times \mathbb{P}(X = x, U = u | V = v)
\end{aligned}$$

The probability of MAI over the entire study given vaccination status v is

$$\begin{aligned}
&\mathbb{P}(\text{person classified as case of MAI} | V = v) \\
&= \sum_{j=1}^{J} \left[\sum_{x=0}^{1} \sum_{u=0}^{1} \delta_{2vu} \gamma_{jvx} \times \mathbb{P}(X = x, U = u | V = v) \right],
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{P}(X = x, U = u | V = v) &= \frac{\mathbb{P}(V = v | X = x, U = u) \mathbb{P}(X = x, U = u)}{\mathbb{P}(V = v)} \\
&= \frac{\mathbb{P}(V = v | X = x, U = u) \mathbb{P}(X = x, U = u)}{\sum_{x=0}^{1} \sum_{u=0}^{1} \mathbb{P}(V = v | X = x, U = u) \mathbb{P}(X = x, U = u)}.
\end{aligned}$$

Using our model parameters we can rewrite $\mathbb{P}(X = x, U = u|V = v)$ as

$$\begin{aligned}
\mathbb{P}(X=0,U=0|V=0) &= \frac{(1-\alpha_{00})\pi_{00}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
\mathbb{P}(X=0,U=1|V=0) &= \frac{(1-\alpha_{01})\pi_{01}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
\mathbb{P}(X=1,U=0|V=0) &= \frac{(1-\alpha_{10})\pi_{10}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
\mathbb{P}(X=1,U=1|V=0) &= \frac{(1-\alpha_{11})\pi_{11}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
\mathbb{P}(X=0,U=0|V=1) &= \frac{\alpha_{00}\pi_{00}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
\mathbb{P}(X=0,U=1|V=1) &= \frac{\alpha_{01}\pi_{01}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
\mathbb{P}(X=1,U=0|V=1) &= \frac{\alpha_{10}\pi_{10}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
\mathbb{P}(X=1,U=1|V=1) &= \frac{\alpha_{11}\pi_{11}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}}
\end{aligned}$$

Again, when evaluating the true VE, we assume random vaccination ($\alpha_{xu} = \alpha$ $\forall x, u$), thus the above expressions of $\mathbb{P}(X = x, U = u|V = v)$ can be simplified as

$$\begin{aligned}
\mathbb{P}(X=0,U=0|V=0) &= \frac{(1-\alpha)\pi_{00}}{(1-\alpha)(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{00} \\
\mathbb{P}(X=0,U=1|V=0) &= \frac{(1-\alpha)\pi_{01}}{(1-\alpha)(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{01} \\
\mathbb{P}(X=1,U=0|V=0) &= \frac{(1-\alpha)\pi_{10}}{(1-\alpha)(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{10} \\
\mathbb{P}(X=1,U=1|V=0) &= \frac{(1-\alpha)\pi_{11}}{(1-\alpha)(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{11} \\
\mathbb{P}(X=0,U=0|V=1) &= \frac{\alpha\pi_{00}}{\alpha(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{00} \\
\mathbb{P}(X=0,U=1|V=1) &= \frac{\alpha\pi_{01}}{\alpha(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{01} \\
\mathbb{P}(X=1,U=0|V=1) &= \frac{\alpha\pi_{10}}{\alpha(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{10} \\
\mathbb{P}(X=1,U=1|V=1) &= \frac{\alpha\pi_{11}}{\alpha(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})} &= \pi_{11}
\end{aligned}$$

Therefore, the true VE against MAI (VE_{TMAI}) is

$$VE_{TMAI} = 1 - \frac{\sum_{j=1}^{J}\left\{\gamma_{j10}\left[\delta_{210}\pi_{00}+\delta_{211}\pi_{01}\right]+\gamma_{j11}\left[\delta_{210}\pi_{10}+\delta_{211}\pi_{11}\right]\right\}}{\sum_{j=1}^{J}\left\{\gamma_{j00}\left[\delta_{200}\pi_{00}+\delta_{201}\pi_{01}\right]+\gamma_{j01}\left[\delta_{200}\pi_{10}+\delta_{201}\pi_{11}\right]\right\}}$$

B.2 Probability of Being a Case

B.2.1 ASC Study

For a person to be considered a case in week j, then they have to be diagnosed with FARI in week j (i.e., $T_j = 1$) and have an ARI in week j (i.e., $Y_j > 0$). Since we do not assume perfect influenza test sensitivity and specificity, a person may also be misdiagnosed with influenza infection.

Consider a person with a given (v, x, u) combination. Let C_j be the event that the person was identified as a case in week j ($j = 1, \ldots, J$), i.e. $C_j = \{T_j = 1\}$. Then the event that the person is a case is the union of all the events C_j, $j = 1, \ldots, J$. We write the event C_j as the union of two disjoint events: $A_j = \{T_j = 1, Y_j = 1\}$ and $B_j = \{T_j = 1, Y_j = 2\}$, where $A_1, \ldots, A_J$ are independent and $B_1, \ldots, B_J$ are mutually exclusive. Let $A = \bigcup_{j=1}^{J} A_j$, $B = \bigcup_{j=1}^{J} B_j$, and $C = \bigcup_{j=1}^{J} C_j$.

We can write the probability of being a case in an ASC study (C) as

$$\mathbb{P}(C) \;=\; \mathbb{P}(A \bigcup B) \;=\; \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \bigcap B)$$

Since $A_1, \ldots, A_J$ are independent

$$\begin{aligned} \mathbb{P}(A) &= 1 - \mathbb{P}(\bigcap \bar{A}) \\ &= 1 - \prod_{j=1}^{J} \mathbb{P}(\bar{A}_j) \\ &= 1 - \prod_{j=1}^{J} (1 - \tau_1 \beta_{jvx}), \end{aligned}$$

where $\bar{A}$ is the event that a person never falsely tested positive for influenza.

Since $B_1, \ldots, B_J$ are mutually exclusive,

$$\begin{aligned}\mathbb{P}(\bigcup_{j=1}^{J} B_j) &= \sum_{j=1}^{J} \mathbb{P}(B_j) \\ &= \sum_{j=1}^{J} (\tau_2 \gamma_{jvx}),\end{aligned}$$

where $\bar{B}$ is the event that a person never correctly tested positive for influenza.

Finally, we can write $\mathbb{P}(A \bigcap B)$ as

$$\begin{aligned}\mathbb{P}(A \bigcap B) &= \sum_{j=1}^{J}\sum_{k=1}^{J} \mathbb{P}(A_j \cap B_k), \text{ where } j \neq k \\ &= \sum_{j=1}^{J}\sum_{k=1}^{J} \mathbb{P}(\{T_j = 1, Y_j = 1\} \cap (\{T_k = 1, Y_k = 2\}), \text{ where } j \neq k \\ &= \sum_{j=1}^{J}\sum_{k=1}^{J} \mathbb{P}(T_j = 1, Y_j = 1)\mathbb{P}(T_k = 1, Y_k = 2e), \text{ where } j \neq k \\ &= \sum_{j=1}^{J}\sum_{k=1}^{J} (\tau_1 \beta_{jvx} \times \tau_2 \gamma_{kvx}), \text{ where } j \neq k\end{aligned}$$

The probability of being a case in an ASC study for a given v can be written as

$$\begin{aligned}\mathbb{P}(C|V = v) &= \sum_{x=0}^{1} \Bigg\{ \left[1 - \prod_{j=1}^{J} (1 - \tau_1 \beta_{jvx}) + \sum_{j=1}^{J} (\tau_2 \gamma_{jvx}) - \sum_{j=1}^{J}\sum_{k=1}^{J} (\tau_1 \beta_{jvx} \times \tau_2 \gamma_{kvx}) \right] \\ &\quad \times \mathbb{P}(X = x|V = v) \Bigg\},\end{aligned}$$

where $j \neq k$.

Thus, the VE estimate from an ASC study is

$$VE_{ASC} = 1 - \frac{\sum_{x=0}^{1}\left\{\left[1 - \prod_{j=1}^{J}(1-\tau_1\beta_{j1x}) + \sum_{j=1}^{J}(\tau_2\gamma_{j1x}) - \sum_{j=1}^{J}\sum_{k=1}^{J}(\tau_1\beta_{j1x}\times\tau_2\gamma_{k1x})\right]\times\mathbb{P}(X=x|V=1)\right\}}{\sum_{x=0}^{1}\left\{\left[1 - \prod_{j=1}^{J}(1-\tau_1\beta_{j0x}) + \sum_{j=1}^{J}(\tau_2\gamma_{j0x}) - \sum_{j=1}^{J}\sum_{k=1}^{J}(\tau_1\beta_{j0x}\times\tau_2\gamma_{k0x})\right]\times\mathbb{P}(X=x|V=0)\right\}}.$$

B.2.2 PSC Study

In a PSC study, a person is considered a case in week j if they have an ARI in week j (i.e., $Y_J > 0$), seek medical care for their ARI in week j (i.e., $M_j = 1$), and test positive for influenza infection (i.e., $T_j = 1$).

Let C_j^* be the event that the person was identified as a case in week j ($j = 1, \ldots, J$), i.e. $C_j^* = \{T_j = 1, M_j = 1\}$. Then the event that the person is a case is the union of all the events C_j^*, $j = 1, \ldots, J$. We can write the event C_j^* as the union of two disjoint events: $A_j^* = \{T_j = 1, M_j = 1, Y_j = 1\}$ and $B_j^* = \{T_j = 1, M_j = 1, Y_j = 2\}$, where $A_1^*, \ldots, A_J^*$ are independent and $B_1^*, \ldots, B_J^*$ are mutually exclusive. Let $A^* = \bigcup_{j=1}^{J} A_j^*$, $B^* = \bigcup_{j=1}^{J} B_j^*$, and $C^* = \bigcup_{j=1}^{J} C_j^*$. As in an ASC study, we can write the probability of being a case in a PSC study (C^*) as $\mathbb{P}(C^*) = \mathbb{P}(A^*) + \mathbb{P}(B^*) - \mathbb{P}(\mathbb{A}^* \bigcap \mathbb{B}^*)$.

Using the parameters from our model, we can write the probability of being an PSC case for a given (v, x, u) as

$$\begin{aligned}
&\mathbb{P}(C^*|V=v, X=x, U=u)\\
&= 1 - \prod_{j=1}^{J}(1-\tau_1\delta_{1vu}\beta_{jvx}) + \sum_{j=1}^{J}(\tau_2\delta_{2vu}\gamma_{jvx}) - \sum_{j=1}^{J}\sum_{k=1}^{J}\left[\tau_1\delta_{1vu}\beta_{jvx}\times\tau_2\delta_{2vu}\gamma_{kvx}\right],
\end{aligned}$$

where $k \neq j$.

$$\begin{aligned}
&\mathbb{P}(C^*|V=v) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1}\left\{1 - \prod_{j=1}^{J}(1-\tau_1\delta_{1vu}\beta_{jvx}) + \sum_{j=1}^{J}(\tau_2\delta_{2vu}\gamma_{jvx}) - \sum_{j=1}^{J}\sum_{k=1}^{J}[\tau_1\delta_{1vu}\beta_{jvx} \times \tau_2\delta_{2vu}\gamma_{kvx}]\right\} \\
&\times \mathbb{P}(X=x, U=u|V=v),
\end{aligned}$$

where $k \neq j$.

Thus, the VE estimate from an PSC study is

$$VE_{PSC} = 1 - \frac{\mathbb{P}(C^*|V=1)}{\mathbb{P}(C^*|V=0)}$$

B.2.3 TN and TCC Studies

The probability of being a case in week j is the probability of being diagnosed with FARI in week j. The probability of having FARI in week j is the same as the probability of having FARI in week j and not having a FARI in weeks 1 to $j-1$. For a given (v, x, u) we can write

$$\mathbb{P}(Y_j = 2|V=v, X=x) = \mathbb{P}(Y_j = 2, \mathbf{Y_{j-1}} < \mathbf{2}|V=v, X=x) = \gamma_{jvx}.$$

Also, when $Y_1, \ldots, Y_{j-1}$ are given, $M_1, \ldots, M_{j-1}$ are independent, thus

$$
\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, V = v, U = u) \\
&= \mathbb{P}(M_1 = 0 | Y_1 < 2, V = v, U = u) \cdots \mathbb{P}(M_{j-1} = 0 | Y_{j-1} < 2, V = v, U = u) \\
&= \prod_{k=1}^{j-1} \mathbb{P}(M_k = 0 | Y_k < 2, V = v, U = u) \\
&= \prod_{k=1}^{j-1} \frac{\mathbb{P}(M_k = 0, Y_k < 2 | V = v, U = u)}{\mathbb{P}(Y_k < 2 | V = v, X = x)} \\
&= \prod_{k=1}^{j-1} \frac{\sum_{y=0}^{1} \mathbb{P}(M_k = 0, Y_k = y | V = v, X = x, U = u)}{\sum_{y=0}^{1} \mathbb{P}(Y_k = y | V = v, X = x)} \\
&= \prod_{k=1}^{j-1} \frac{\sum_{y=0}^{1} \mathbb{P}(M_k = 0 | Y_k = y, V = v, U = u) \mathbb{P}(Y_k = y | V = v, X = x)}{\sum_{y=0}^{1} \mathbb{P}(Y_k = y | V = v, X = x)} \\
&= \prod_{k=1}^{j-1} \frac{1 \times (1 - \beta_{kvx} - \gamma_{kvx}) + (1 - \delta_{1vu}) \beta_{kvx}}{(1 - \gamma_{kvx})} \\
&= \prod_{k=1}^{j-1} \frac{1 - \gamma_{kvx} - \delta_{1vu} \beta_{kvx}}{(1 - \gamma_{kvx})}
\end{aligned}
$$

Therefore, the probability of being considered a case in week j for a given (v, x, u) is:

$$\mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0}, T_j = 1 | V = v, X = x, U = u)$$
$$= \sum_{y=1}^{2} \mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0}, T_j = 1 | Y_j = y, V = v, U = u) \mathbb{P}(Y_j = y | V = v, X = x)$$

where

$$\mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0}, T_j = 1 | Y_j = y, V = v, U = u)$$
$$= \mathbb{P}(T_j = 1 | Y_j = y) \mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | Y_j = y, V = v, U = u)$$

and

$$\mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | Y_j = y, V = v, U = u)$$
$$= \mathbb{P}(M_j = 1 | Y_j = y, V = v, U = u) \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, V = v, U = u).$$

The probability of being considered a case in week j for a given v is:

$$
\begin{aligned}
&\mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0}, T_j = 1 | V = 1) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1}\left\{\left[\tau_2\delta_{21u}\gamma_{j1x}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k1x}-\delta_{11u}\beta_{k1x}}{1-\gamma_{k1x}}\right) + \tau_1\delta_{11u}\beta_{j1x}\prod_{k=1}^{j-1}(1-\delta_{11u}\beta_{k1x}-\delta_{21u}\gamma_{k1x})\right]\right. \\
&\times \left.\mathbb{P}(X=x, U=u|V=1)\right\} \\
&= \left[\delta_{210}\gamma_{j10}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k10}-\delta_{110}\beta_{k10}}{1-\gamma_{k10}}\right) + \tau_1\delta_{110}\beta_{j10}\prod_{k=1}^{j-1}(1-\delta_{110}\beta_{k10}-\delta_{210}\gamma_{k10})\right] \\
&\times\frac{\alpha_{00}\pi_{00}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
&+ \left[\delta_{211}\gamma_{j10}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k10}-\delta_{111}\beta_{k10}}{1-\gamma_{k10}}\right) + \tau_1\delta_{111}\beta_{j10}\prod_{k=1}^{j-1}(1-\delta_{111}\beta_{k10}-\delta_{211}\gamma_{k10})\right] \\
&\times\frac{\alpha_{01}\pi_{01}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
&+ \left[\delta_{210}\gamma_{j11}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k11}-\delta_{110}\beta_{k11}}{1-\gamma_{k11}}\right) + \tau_1\delta_{110}\beta_{j11}\prod_{k=1}^{j-1}(1-\delta_{110}\beta_{k11}-\delta_{210}\gamma_{k11})\right] \\
&\times\frac{\alpha_{10}\pi_{10}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
&+ \left[\delta_{211}\gamma_{j11}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k11}-\delta_{111}\beta_{k11}}{1-\gamma_{k11}}\right) + \tau_1\delta_{111}\beta_{j11}\prod_{k=1}^{j-1}(1-\delta_{111}\beta_{k11}-\delta_{211}\gamma_{k11})\right] \\
&\times\frac{\alpha_{11}\pi_{11}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0}, T_j = 1 | V = 0) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1}\left\{\left[\tau_2\delta_{20u}\gamma_{j0x}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k0x}-\delta_{10u}\beta_{k0x}}{1-\gamma_{k0x}}\right) + \tau_1\delta_{10u}\beta_{j0x}\prod_{k=1}^{j-1}(1-\delta_{10u}\beta_{k0x}-\delta_{20u}\gamma_{k0x})\right]\right. \\
&\times \left.\mathbb{P}(X=x, U=u|V=0)\right\} \\
&= \left[\delta_{200}\gamma_{j00}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k00}-\delta_{100}\beta_{k00}}{1-\gamma_{k00}}\right) + \tau_1\delta_{00u}\beta_{j00}\prod_{k=1}^{j-1}(1-\delta_{100}\beta_{k00}-\delta_{200}\gamma_{k00})\right] \\
&\times\frac{(1-\alpha_{00})\pi_{00}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
&+ \left[\delta_{201}\gamma_{j00}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k00}-\delta_{101}\beta_{k00}}{1-\gamma_{k00}}\right) + \tau_1\delta_{101}\beta_{j00}\prod_{k=1}^{j-1}(1-\delta_{101}\beta_{k00}-\delta_{201}\gamma_{k00})\right] \\
&\times\frac{(1-\alpha_{01})\pi_{01}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
&+ \left[\delta_{200}\gamma_{j01}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k01}-\delta_{100}\beta_{k01}}{1-\gamma_{k01}}\right) + \tau_1\delta_{100}\beta_{j01}\prod_{k=1}^{j-1}(1-\delta_{100}\beta_{k01}-\delta_{200}\gamma_{k01})\right] \\
&\times\frac{(1-\alpha_{10})\pi_{10}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
&+ \left[\delta_{201}\gamma_{j01}\prod_{k=1}^{j-1}\left(\frac{1-\gamma_{k01}-\delta_{101}\beta_{k01}}{1-\gamma_{k01}}\right) + \tau_1\delta_{101}\beta_{j01}\prod_{k=1}^{j-1}(1-\delta_{101}\beta_{k01}-\delta_{201}\gamma_{k01})\right] \\
&\times\frac{(1-\alpha_{11})\pi_{11}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}}
\end{aligned}
$$

B.3 Probability of Being a TN Control

For a given (v, x, u) the probability of being considered a TN control in week j can be expressed as:

$$
\begin{aligned}
&\mathbb{P}(Y_j = 1, M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | V = v, X = x, U = u) \\
&= \mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | Y_j = 1, V = v, X = x, U = u)\mathbb{P}(Y_j = 1 | V = v, X = x)
\end{aligned}
$$

The second term in the above expression is simply β_{jvx}, while the first term can be further written as:

$$\begin{aligned}
&\mathbb{P}(M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | Y_j = 1, V = v, X = x, U = u) \\
&= \mathbb{P}(M_j = 1 | Y_j = 1, V = v, X = x, U = u) \times \prod_{k=1}^{j-1} \mathbb{P}(M_k = 0 | V = v, X = x, U = u) \\
&= \delta_{1vu} \times \prod_{k=1}^{j-1} \mathbb{P}(M_k = 0 | V = v, X = x, U = u),
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(M_k = 0 | V = v, X = x, U = u) \\
&= \sum_{y=0}^{2} \mathbb{P}(M_k = 0 | Y_k = y, V = v, X = x, U = u) \mathbb{P}(Y_k = y | V = v, X = x) \\
&= (1 - \beta_{kvx} - \gamma_{kvx}) + (1 - \delta_{1vu})\beta_{kvx} + (1 - \delta_{2vu})\gamma_{kvx} \\
&= 1 - \delta_{1vu}\beta_{kvx} - \delta_{2vu}\gamma_{kvx}.
\end{aligned}$$

Thus, for a given (v, x, u) the probability of being considered a TN control in week j is

$$\begin{aligned}
&\mathbb{P}(Y_j = 1, M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | V = v, X = x, U = u) \\
&= \delta_{1vu}\beta_{jvx} \prod_{k=1}^{j-1} (1 - \delta_{1vu}\beta_{kvx} - \delta_{2vu}\gamma_{kvx}).
\end{aligned}$$

The probability of being considered a TN control in week j for a given v is:

$$
\begin{aligned}
&\mathbb{P}(Y_j = 1, M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | V = 1) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1}\left\{\delta_{11u}\beta_{j1x}\prod_{k=1}^{j-1}(1-\delta_{11u}\beta_{k1x}-\delta_{21u}\gamma_{k1x}) \times \mathbb{P}(X=x, U=u|V=1)\right\} \\
&= \delta_{110}\beta_{j10}\prod_{k=1}^{j-1}(1-\delta_{110}\beta_{k10}-\delta_{210}\gamma_{k10}) \times \frac{\alpha_{00}\pi_{00}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
&+\delta_{111}\beta_{j10}\prod_{k=1}^{j-1}(1-\delta_{111}\beta_{k10}-\delta_{211}\gamma_{k10}) \times \frac{\alpha_{01}\pi_{01}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
&+\delta_{110}\beta_{j11}\prod_{k=1}^{j-1}(1-\delta_{110}\beta_{k11}-\delta_{210}\gamma_{k11}) \times \frac{\alpha_{10}\pi_{10}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}} \\
&+\delta_{111}\beta_{j11}\prod_{k=1}^{j-1}(1-\delta_{111}\beta_{k11}-\delta_{211}\gamma_{k11}) \times \frac{\alpha_{11}\pi_{11}}{\alpha_{00}\pi_{00}+\alpha_{01}\pi_{01}+\alpha_{10}\pi_{10}+\alpha_{11}\pi_{11}}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{P}(Y_j = 1, M_j = 1, \mathbf{M_{j-1}} = \mathbf{0} | V = 0) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1}\left\{\delta_{10u}\beta_{j0x}\prod_{k=1}^{j-1}(1-\delta_{10u}\beta_{k0x}-\delta_{20u}\gamma_{k0x}) \times \mathbb{P}(X=x, U=u|V=0)\right\} \\
&= \delta_{100}\beta_{j00}\prod_{k=1}^{j-1}(1-\delta_{100}\beta_{k00}-\delta_{200}\gamma_{k00}) \times \frac{(1-\alpha_{00})\pi_{00}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
&+\delta_{101}\beta_{j00}\prod_{k=1}^{j-1}(1-\delta_{101}\beta_{k00}-\delta_{201}\gamma_{k00}) \times \frac{(1-\alpha_{01})\pi_{01}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
&+\delta_{100}\beta_{j01}\prod_{k=1}^{j-1}(1-\delta_{100}\beta_{k01}-\delta_{201}\gamma_{k01}) \times \frac{(1-\alpha_{10})\pi_{10}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}} \\
&+\delta_{101}\beta_{j01}\prod_{k=1}^{j-1}(1-\delta_{101}\beta_{k01}-\delta_{201}\gamma_{k01}) \times \frac{(1-\alpha_{11})\pi_{11}}{(1-\alpha_{00})\pi_{00}+(1-\alpha_{01})\pi_{01}+(1-\alpha_{10})\pi_{10}+(1-\alpha_{11})\pi_{11}}
\end{aligned}
$$

To obtain the final 2x2 table for a TN study, we collapse across X and U for each week j and then sum over all weeks $1, \ldots, J$. Finally, we can obtain the VE estimate from a TN study using the OR from the final 2x2 table. The VE

estimate from a TN study is

$$VE_{TN} = 1 - OR_{TN} = \frac{a \times d_{TN}}{b \times c_{TN}},$$

where

$$\begin{aligned}
a &= \sum_{j=1}^{J}\left\{\sum_{x=0}^{1}\sum_{u=0}^{1}\left[\delta_{21u}\gamma_{j1x}\prod_{k=1}^{j-1}\left[\frac{1-\gamma_{k1x}-\delta_{11u}\beta_{k1x}}{1-\gamma_{k1x}}\right]\times\mathbb{P}(X=x,U=u|V=1)\right]\right\}\\
b &= \sum_{j=1}^{J}\left\{\sum_{x=0}^{1}\sum_{u=0}^{1}\left[\delta_{20u}\gamma_{j0x}\prod_{k=1}^{j-1}\left[\frac{1-\gamma_{k0x}-\delta_{10u}\beta_{k0x}}{1-\gamma_{k0x}}\right]\times\mathbb{P}(X=x,U=u|V=0)\right]\right\}\\
c_{TN} &= \sum_{j=1}^{J}\left\{\sum_{x=0}^{1}\sum_{u=0}^{1}\left[\delta_{11u}\beta_{j1x}\prod_{k=1}^{j-1}(1-\delta_{11u}\beta_{k1x}-\delta_{21u}\gamma_{k1x})\times\mathbb{P}(X=x,U=u|V=1)\right]\right\}\\
d_{TN} &= \sum_{j=1}^{J}\left\{\sum_{x=0}^{1}\sum_{u=0}^{1}\left[\delta_{10u}\beta_{j0x}\prod_{k=1}^{j-1}(1-\delta_{10u}\beta_{k0x}-\delta_{20u}\gamma_{k0x})\times\mathbb{P}(X=x,U=u|V=0)\right]\right\}
\end{aligned}$$

B.4 Probability of no ARI

To determine the probability of being considered a TCC control, we need to determine the probability that a person did not have either a FARI or NFARI during the study. Because Y_{Fj} is independent of Y_{Nj} for all j, we can determine the probability of not having a FARI separately from the probability of not having a NFARI.

Due to the independence of the Y_{Nj}s, the joint probability of not having a NFARI in weeks $1,\ldots,J$ is simply the product of the marginal probabilities of not having a NFARI in each week. Thus, for a given (v,x)

$$\begin{aligned}
\mathbb{P}(Y_{N1}=0,\ldots,Y_{NJ}=0|V=v,X=x) &= \prod_{k=1}^{J}\mathbb{P}(Y_{Nk}=0|V=v,X=x)\\
&= \prod_{k=1}^{J}(1-\beta_{kvx}).
\end{aligned}$$

Due to our model assumption that a person can only be infected with influenza once during the season, and thus, only have one FARI, we cannot treat the Y_{Fj}s as independent. To determine the joint probability of not having a FARI in weeks $1, \ldots, J$, we will start with a simple case. For given (v, x) first, consider $J = 2$.

$$\begin{aligned}
\mathbb{P}(Y_{F2} = 0 | Y_{F1} = 0, V = v, X = x) &= 1 - \mathbb{P}(Y_{F2} = 1 | Y_{F1} = 0, V = v, X = x) \\
&= 1 - \frac{\mathbb{P}(Y_{F1} = 0, Y_{F2} = 1 | V = v, X = x)}{\mathbb{P}(Y_{F1} = 0 | V = v, X = x)} \\
&= 1 - \frac{\gamma_{2vx}}{1 - \gamma_{1vx}} \\
&= \frac{1 - \gamma_{1vx} - \gamma_{2vx}}{1 - \gamma_{1vx}}.
\end{aligned}$$

So,

$$\begin{aligned}
&\mathbb{P}(Y_{F1} = 0, Y_{F2} = 0 | V = v, X = x) \\
&= \mathbb{P}(Y_{F2} = 0 | Y_{F1} = 0, V = v, X = x) \mathbb{P}(Y_{F1} = 0 | V = v, X = x) \\
&= \frac{1 - \gamma_{1vx} - \gamma_{2vx}}{1 - \gamma_{1vx}} (1 - \gamma_{1vx}) \\
&= 1 - \gamma_{1vx} - \gamma_{2vx}.
\end{aligned}$$

Now, consider $J = 3$.

$$\begin{aligned}
&\mathbb{P}(Y_{F3} = 0 | Y_{F1} = 0, Y_{F2} = 0, V = v, X = x) \\
&= 1 - \mathbb{P}(Y_{F3} = 0 | Y_{F1} = 0, Y_{F2} = 0, V = v, X = x) \\
&= 1 - \left[\frac{\mathbb{P}(Y_{F1} = 0, Y_{F2} = 0, Y_{F3} = 0 | V = v, X = x)}{\mathbb{P}(Y_{F1} = 0, Y_{F2} = 0 | V = v, X = x)} \right] \\
&= 1 - \frac{\gamma_{3vx}}{1 - \gamma_{1vx} - \gamma_{2vx}} \\
&= \frac{1 - \gamma_{1vx} - \gamma_{2vx} - \gamma_{3vx}}{1 - \gamma_{1vx} - \gamma_{2vx}}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}(Y_{F1}=0, Y_{F2}=0, Y_{F3}=0|V=v, X=x)\\
&= \mathbb{P}(Y_{F3}=0|Y_{F1}=0, Y_{F2}=0, V=v, X=x)\mathbb{P}(Y_{F1}=0, Y_{F2}=0|V=v, X=x)\\
&= \frac{1-\gamma_{1vx}-\gamma_{2vx}-\gamma_{3vx}}{1-\gamma_{1vx}-\gamma_{2vx}}(1-\gamma_{1vx}-\gamma_{2vx})\\
&= 1-\gamma_{1vx}-\gamma_{2vx}-\gamma_{3vx}.
\end{aligned}$$

For any j, $j = 1, \ldots, J$,

$$\mathbb{P}(Y_{Fj}=0|Y_{F1}=0, \ldots, Y_{F(j-1)}=0, V=v, X=x) \;=\; \frac{1-\sum_{k=1}^{j}\gamma_{kvx}}{1-\sum_{k=0}^{j-1}\gamma_{kvx}}$$

Thus, the probability that a person does not have an ARI in weeks $1, \ldots, J$ is

$$\mathbb{P}(Y_{F1}=0, \ldots, Y_{FJ}=0|V=v, X=x) \;=\; 1-\sum_{k=1}^{J}\gamma_{kvx},$$

where $\sum_{k=1}^{J}\gamma_{kvx} \leq 1$.

B.4.1 Probability of being a TCC control

For a person to be considered a TCC control they did not have an ARI during the study (i.e., $\mathbf{Y_J} = \mathbf{0}$). Using the results from above, the probability of being

a TCC control for a given (v, x) can be written as

$$\begin{aligned}\mathbb{P}(\mathbf{Y_J} = \mathbf{0}|V = v, X = x) &= \mathbb{P}(Y_{F1} = 0, \ldots, Y_{FJ} = 0|V = v, X = x) \\ &\times \mathbb{P}(Y_{N1} = 0, \ldots, Y_{NJ} = 0|V = v, X = x) \\ &= \left[1 - \sum_{k=1}^{J} \gamma_{kvx}\right] \left[\prod_{k=1}^{J}(1 - \beta_{kvx})\right].\end{aligned}$$

The probability of being a TCC control for a given v is:

$$\begin{aligned}&\mathbb{P}(\mathbf{Y_J} = \mathbf{0}|V = 1) = \sum_{x=0}^{1} \left[1 - \sum_{k=1}^{J} \gamma_{k1x}\right] \left[\prod_{k=1}^{J}(1 - \beta_{k1x})\right] \times \mathbb{P}(X = x|V = 1) \\ &= \left[1 - \sum_{k=1}^{J} \gamma_{k10}\right] \left[\prod_{k=1}^{J}(1 - \beta_{k10})\right] \times \frac{\alpha_{00}\pi_{00} + \alpha_{01}\pi_{01}}{\alpha_{00}\pi_{00} + \alpha_{01}\pi_{01} + \alpha_{10}\pi_{10} + \alpha_{11}\pi_{11}} \\ &+ \left[1 - \sum_{k=1}^{J} \gamma_{k11}\right] \left[\prod_{k=1}^{J}(1 - \beta_{k11})\right] \times \frac{\alpha_{10}\pi_{10} + \alpha_{11}\pi_{11}}{\alpha_{00}\pi_{00} + \alpha_{01}\pi_{01} + \alpha_{10}\pi_{10} + \alpha_{11}\pi_{11}}\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}(\mathbf{Y_J} = \mathbf{0}|V = 0) = \sum_{x=0}^{1} \left[1 - \sum_{k=1}^{J} \gamma_{k0x}\right] \left[\prod_{k=1}^{J}(1 - \beta_{k0x})\right] \times \mathbb{P}(X = x|V = 0) \\ &= \left[1 - \sum_{k=1}^{J} \gamma_{k00}\right] \left[\prod_{k=1}^{J}(1 - \beta_{k00})\right] \times \frac{(1 - \alpha_{00})\pi_{00} + (1 - \alpha_{01})\pi_{01}}{(1 - \alpha_{00})\pi_{00} + (1 - \alpha_{01})\pi_{01} + (1 - \alpha_{10})\pi_{10} + (1 - \alpha_{11})\pi_{11}} \\ &+ \left[1 - \sum_{k=1}^{J} \gamma_{k01}\right] \left[\prod_{k=1}^{J}(1 - \beta_{k01})\right] \times \frac{(1 - \alpha_{10})\pi_{10} + (1 - \alpha_{11})\pi_{11}}{(1 - \alpha_{00})\pi_{00} + (1 - \alpha_{01})\pi_{01} + (1 - \alpha_{10})\pi_{10} + (1 - \alpha_{11})\pi_{11}}.\end{aligned}$$

Finally, we can obtain the VE estimate from a TCC study using the OR from the final 2x2 table. The VE estimate from a TCC study is

$$VE_{TCC} = 1 - OR_{TCC} = \frac{a \times d_{TCC}}{b \times c_{TCC}},$$

where

$$
\begin{aligned}
a &= \sum_{j=1}^{J}\left\{\sum_{x=0}^{1}\sum_{u=0}^{1}\left[\delta_{21u}\gamma_{j1x}\prod_{k=1}^{j-1}\left[\frac{1-\gamma_{k1x}-\delta_{11u}\beta_{k1x}}{1-\gamma_{k1x}}\right]\times\mathbb{P}(X=x,U=u|V=0)\right]\right\}\\
b &= \sum_{j=1}^{J}\left\{\sum_{x=0}^{1}\sum_{u=0}^{1}\left[\delta_{20u}\gamma_{j0x}\prod_{k=1}^{j-1}\left[\frac{1-\gamma_{k0x}-\delta_{10u}\beta_{k0x}}{1-\gamma_{k0x}}\right]\times\mathbb{P}(X=x,U=u|V=0)\right]\right\}\\
c_{TCC} &= \left\{\sum_{x=0}^{1}\left[1-\sum_{k=1}^{J}\gamma_{k0x}\right]\left[\prod_{k=1}^{J}(1-\beta_{k0x})\right]\times\mathbb{P}(X=x|V=1)\right\}\\
d_{TCC} &= \left\{\sum_{x=0}^{1}\left[1-\sum_{k=1}^{J}\gamma_{k0x}\right]\left[\prod_{k=1}^{J}(1-\beta_{k0x})\right]\times\mathbb{P}(X=x|V=0)\right\}
\end{aligned}
$$

Note, a and b are the same for TN and TCC studies.

B.5 Input Parameters

Table B1 shows the parameter input values used in the calculations and simulations. Weekly values of the probabilities of NFARI (βs) and FARI (γs) are based on the prevalence of illness from surveillance data collected by the CDC. Parameters whose input values are a range were varied by 0.1 for each calculation scenario (see section 3.2.5).

Table B1: Input parameter values used in calculations and simulations.

Parameters	Input Values	Parameters	Input Values
π_{11}	0.4	γ_{101}	0.00103726
π_{10}	0.4	γ_{201}	0.0015696
π_{01}	0.1	γ_{301}	0.0018823
π_{11}	0.1	γ_{401}	0.00210545
α_{11}	0.6	γ_{501}	0.00237509
α_{10}	0.3	γ_{601}	0.0030613
α_{01}	0.9	γ_{701}	0.00394527
α_{00}	0.45	γ_{801}	0.00507414
β_{101}	0.020795	γ_{901}	0.006394935
β_{201}	0.02071	γ_{1001}	0.007294342
β_{301}	0.020814	γ_{1101}	0.007909946
β_{401}	0.020773	γ_{1201}	0.008418902
β_{501}	0.024733	γ_{1301}	0.008899444
β_{601}	0.027691	γ_{1401}	0.007516386
β_{701}	0.027538	γ_{1501}	0.005374278
β_{801}	0.027601	γ_{1601}	0.003840503
β_{901}	0.028493	γ_{1701}	0.002726108
β_{1001}	0.034245	γ_{1801}	0.000629997
β_{1101}	0.034375	θ_{γ}	0.563
β_{1201}	0.03441	ϕ_{γ}	1.0-2.0
β_{1301}	0.0333	δ_{101}	0.25
β_{1401}	0.027464	θ_{δ_1}	1.0
β_{1501}	0.027515	μ_{δ_1}	0.5-1.0
β_{1601}	0.027607	δ_{201}	0.40
β_{1701}	0.027593	θ_{δ_2}	0.5-1.0
β_{1801}	0.007901	μ_{δ_2}	0.5-1.0
θ_{β}	0.5-2.0	τ_1	0
ϕ_{β}	1.0-2.0	τ_2	1

We assume $\mu_{\delta_1} = \mu_{\delta_2}$.

Appendix C

Appendix to Chapter 4

C.1 Vaccination Over Time

C.1.1 True VE

True VE Against SI

A person is considered a true case of SI if s/he develops an influenza ARI during the study $(Y_j = 2)$ for one j, $j = j_v, \ldots, J$, where j_v is the first week of vaccination. Let

$$\gamma_{jv} = \mathbb{P}(Y_j = 2|V_j = v) = \frac{C_{jv}}{N_{jv}}$$

and

$$\eta_j = \mathbb{P}(V_j = 1) = \sum_{h=1}^{j} \alpha_h = \frac{N_{j1}}{N}$$

Then, we can write

$$\begin{aligned} N_{j1} &= N \cdot \eta_j \\ C_{j1} &= N \cdot \eta_j \cdot \gamma_{j1} \\ C_{j0} &= N \cdot (1 - \eta_j) \cdot \gamma_{j0} \end{aligned}$$

We can also write $\mathbb{P}(Y_j = 2|V_j = v)$ in terms of our covariates X and U as follows:

$$\begin{aligned}
\gamma_{jv} &= \sum_{x=0}^{1}\sum_{u=0}^{1} \mathbb{P}(Y_j = 2|V_j = v, X = x, U = u) \times \mathbb{P}(X = x, U = u|V_j = v) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{jvx} \times \frac{\mathbb{P}(V_j = v|X = x, U = u) \times \mathbb{P}(X = x, U = u)}{\mathbb{P}(V_j = v)} \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{jvx}\pi_{xu}.
\end{aligned}$$

Thus, for $v = 1$,

$$\begin{aligned}
\gamma_{j1x} &= \sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j1x}\pi_{xu} \\
\Rightarrow \tfrac{C_{j1}}{N_{j1}} &= \sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j1x}\pi_{xu} \\
\Rightarrow C_{j1} &= N_{j1}\sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j1x}\pi_{xu} \\
\Rightarrow C_{j1} &= N\eta_j\sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j1x}\pi_{xu}
\end{aligned}$$

Similarly, $C_{j0} = N(1-\eta_j)\sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j0x}\pi_{xu}$.

Therefore,

$$\frac{\sum_{j=j_v}^{J} C_{j1}}{\sum_{j=j_v}^{J} N_{j1}} = \frac{\sum_{j=j_v}^{J}\left[\eta_j\sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j1x}\pi_{xu}\right]}{\sum_{j=j_v}^{J}\eta_j}$$

and

$$\frac{\sum_{j=j_v}^{J} C_{j0}}{\sum_{j=j_v}^{J} N_{j0}} = \frac{\sum_{j=j_v}^{J}\left[(1-\eta_j)\sum_{x=0}^{1}\sum_{u=0}^{1} \gamma_{j1x}\pi_{xu}\right]}{\sum_{j=j_v}^{J}(1-\eta_j)}.$$

Finally,

$$VE_{TSI} = 1 - \frac{\sum_{j=j_v}^{J} C_{j1} / \sum_{j=j_v}^{J} N_{j1}}{\sum_{j=j_v}^{J} C_{j0} / \sum_{j=j_v}^{J} N_{j0}}$$

True VE Against MAI

A person is considered a true case of MAI if s/he develops an influenza ARI during the study and seeks medical care for this ARI ($Y_j = 2, M_j = 1$) for one j, $j = j_v, \ldots, J$, where j_v is the first week of vaccination.

$$\begin{aligned}
\frac{C^*_{jv}}{N_{jv}} &= \mathbb{P}(Y_j = 2, M_j = 1 | V_j = v) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1} \mathbb{P}(Y_j = 2, M_j = 1 | V_j = v, X = x, U = u) \times \mathbb{P}(X = x, U = u | V_j = v) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1} [\mathbb{P}(M_j = 1 | Y_j = 2, V_j = v, X = x, U = u) \times \mathbb{P}(Y_j = 2 | V_j = v, X = x, U = u) \\
&\times \mathbb{P}(X = x, U = u | V_j = v)]
\end{aligned}$$

Thus,

$$\begin{aligned}
C^*_{j1} &= N\eta_j \sum_{x=0}^{1}\sum_{u=0}^{1} \delta_{21u}\gamma_{j1x}\pi_{xu} \\
C^*_{j0} &= N(1-\eta_j) \sum_{x=0}^{1}\sum_{u=0}^{1} \delta_{20u}\gamma_{j0x}\pi_{xu}.
\end{aligned}$$

Therefore,

$$\frac{\sum_{j=j_v}^{J} C^*_{j1}}{\sum_{j=j_v}^{J} N_{j1}} = \frac{\sum_{j=j_v}^{J} \left[\eta_j \sum_{x=0}^{1}\sum_{u=0}^{1} \delta_{21u}\gamma_{j1x}\pi_{xu}\right]}{\sum_{j=j_v}^{J} \eta_j}$$

and

$$\frac{\sum_{j=j_v}^{J} C^*_{j0}}{\sum_{j=j_v}^{J} N_{j0}} = \frac{\sum_{j=j_v}^{J}\left[(1-\eta_j)\sum_{x=0}^{1}\sum_{u=0}^{1}\delta_{20u}\gamma_{j0x}\pi_{xu}\right]}{\sum_{j=j_v}^{J}(1-\eta_j)}.$$

Finally,

$$VE_{TMAI} = 1 - \frac{\sum_{j=j_v}^{J} C^*_{j1} / \sum_{j=j_v}^{J} N_{j1}}{\sum_{j=j_v}^{J} C^*_{j0} / \sum_{j=j_v}^{J} N_{j0}}$$

C.1.2 VE Estimates from TN Studies

Probability of Being a Case

A person is considered a case in week j, if:

- they did not seek medical care for any ARI prior to week j, so $M_k = 0$ for every week $k = 1, \ldots, j-1$ (i.e., $\mathbf{M_{j-1}} = \mathbf{0}$)
- they seek medical care for their ARI in week j (i.e., $M_j = 1$)
- they have FARI in week j (i.e., $Y_j = 2$).

We assume perfect influenza test sensitivity and specificity. We can write the probabilities of being a case in week j and having a particular vaccination status as:

$$\begin{aligned}
\mathbb{P}(\text{vaccinated case in week } j) &= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K \leq j) \\
\mathbb{P}(\text{unvaccinated case in week } j) &= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K > j).
\end{aligned}$$

We can write the probability of being classified as a vaccinated case in week j for a given (x, u) as

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K \leq j | X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2 | K \leq j, X = x, U = u) \mathbb{P}(K \leq j | X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, K \leq j, X = x, U = u) \\
&\times \mathbb{P}(Y_j = 2 | K \leq j, X = x, U = u) \mathbb{P}(K \leq j | X = x, U = u),
\end{aligned}$$

and the probability of being classified as an unvaccinated case in week j for a given (x, u) as

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K > j | X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, K > j, X = x, U = u) \\
&\times \mathbb{P}(Y_j = 2 | K > j, X = x, U = u) \mathbb{P}(K > j | X = x, U = u).
\end{aligned}$$

Under our assumption that a person can have only one FARI during the study

$$\begin{aligned}
\mathbb{P}(Y_j = 2 | K \leq j, X = x, U = u) &= \mathbb{P}(Y_j = 2, \mathbf{Y_{j-1}} < \mathbf{2} | K \leq j, X = x, U = u) = \gamma_{j1x} \\
\mathbb{P}(Y_j = 2 | K > j, X = x, U = u) &= \mathbb{P}(Y_j = 2, \mathbf{Y_{j-1}} < \mathbf{2} | K > j, X = x, U = u) = \gamma_{j0x}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, K \leq j, X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, \mathbf{Y_{j-1}} < \mathbf{2}, K \leq j, X = x, U = u) \\
&= \mathbb{P}(M_j = 1 | Y_j = 2, K \leq j, X = x, U = u) \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, K \leq j, X = x, U = u)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, K > j, X = x, U = u) \\
&= \mathbb{P}(M_j = 1 | Y_j = 2, K > j, X = x, U = u) \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, K > j, X = x, U = u)
\end{aligned}$$

where

$$\mathbb{P}(M_j = 1 | Y_j = 2, K \leq j, X = x, U = u) = \delta_{21u}$$

$$\mathbb{P}(M_j = 1 | Y_j = 2, K > j, X = x, U = u) = \delta_{20u}$$

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, K \leq j, X = x, U = u)$$
$$= \prod_{h=j_v}^{j-1} \mathbb{P}(M_h = 0 | Y_h < 2, K \leq j, X = x, U = u),$$

since $M_1, \ldots, M_{j-1}$ are independent given $Y_1, \ldots, Y_{j-1},$ and

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, K > j, X = x, U = u)$$
$$= \prod_{h=j_v}^{j-1} \mathbb{P}(M_h = 0 | Y_h < 2, K > j, X = x, U = u).$$

The probability of not seeking care in week h $(M_h = 0)$ given a vaccinated person does not have FARI $(Y_h < 2)$ can be derived as follows:

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h < 2, K \leq j, X = x, U = u) \\
&= \frac{\mathbb{P}(M_h = 0, Y_h < 2, K \leq j | X = x, U = u)}{\mathbb{P}(Y_h < 2, K \leq j | X = x, U = u)} \\
&= \frac{\sum_{y=0}^{1} \sum_{g=1}^{j-1} \mathbb{P}(M_h = 0, Y_h = y, K = g | X = x, U = u)}{\sum_{y=0}^{1} \sum_{g=1}^{j-1} \mathbb{P}(Y_h = y, K = g | X = x, U = u)}
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(M_h = 0, Y_h = y, K = g | X = x, U = u) \\
&= \mathbb{P}(M_h = 0 | Y_h = y, K = g, X = x, U = u) \mathbb{P}(Y_h = y | K = g, X = x, U = u) \\
&\times \mathbb{P}(K = g | X = x, U = u)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h = 0, K = g, X = x, U = u) = 1 \\
&\mathbb{P}(M_h = 0 | Y_h = 1, K = g, X = x, U = u) = 1 - \delta_{11u} \qquad \text{when } g \leq h \\
&\mathbb{P}(M_h = 0 | Y_h = 1, K = g, X = x, U = u) = 1 - \delta_{10u} \qquad \text{when } g > h \\
&\mathbb{P}(Y_h = 0 | K = g, X = x, U = u) = 1 - \beta_{h1x} - \gamma_{h1x} \qquad \text{when } g \leq h \\
&\mathbb{P}(Y_h = 0 | K = g, X = x, U = u) = 1 - \beta_{h0x} - \gamma_{h0x} \qquad \text{when } g > h \\
&\mathbb{P}(Y_h = 1 | K = g, X = x, U = u) = \beta_{h1x} \qquad \text{when } g \leq h \\
&\mathbb{P}(Y_h = 1 | K = g, X = x, U = u) = \beta_{h0x} \qquad \text{when } g > h \\
&\mathbb{P}(K = g | X = x, U = u) = \alpha_{gxu}
\end{aligned}$$

The probability of not seeking care in week h $(M_h = 0)$ given an unvaccinated person does not have FARI $(Y_h < 2)$ can be derived similarly:

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h < 2, K > j, X = x, U = u) \\
&= \frac{\mathbb{P}(M_h = 0, Y_h < 2, K > j | X = x, U = u)}{\mathbb{P}(Y_h < 2, K > j | X = x, U = u)} \\
&= \frac{\sum_{y=0}^{1} \mathbb{P}(M_h = 0, Y_h = y, K > j | X = x, U = u)}{\sum_{y=0}^{1} \mathbb{P}(Y_h = y, K > j | X = x, U = u)},
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(M_h = 0, Y_h = y, K > j | X = x, U = u) \\
&= \mathbb{P}(M_h = 0 | Y_h = y, K > j, X = x, U = u) \mathbb{P}(Y_h = y | K > j, X = x, U = u) \\
&\times \mathbb{P}(K > j | X = x, U = u)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h = 0, K > j, X = x, U = u) = 1 \\
&\mathbb{P}(M_h = 0 | Y_h = 1, K > j, X = x, U = u) = 1 - \delta_{10u} \\
&\mathbb{P}(Y_h = 0 | K > j, X = x, U = u) = 1 - \beta_{h0x} - \gamma_{h0x} \\
&\mathbb{P}(Y_h = 1 | K > j, X = x, U = u) = \beta_{h0x} \\
&\mathbb{P}(K > j | X = x, U = u) = \sum_{g=j+1}^{J+1} \alpha_{gxu}
\end{aligned}$$

Finally, the overall probabilities of being a vaccinated and an unvaccinated case, respectively, in week j are

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K \leq j)$$
$$= \sum_{x=0}^{1} \sum_{u=0}^{1} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K \leq j, X = x, U = u)\mathbb{P}(X = x, U = u)$$

or

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K > j)$$
$$= \sum_{x=0}^{1} \sum_{u=0}^{1} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K > j, X = x, U = u)\mathbb{P}(X = x, U = u)$$

Probability of Being a TN Control

A person is considered a TN control in week j, if:

- they did not seek medical care for an ARI prior to week j (i.e., $\mathbf{M_{j-1}} = \mathbf{0}$)
- they seek medical care for their ARI in week j (i.e., $M_j = 1$)
- they have NFARI in week j (i.e., $Y_j = 1$)

We can write the probabilities of being a TN control in week j and having a particular vaccination status as:

$$\begin{aligned} \mathbb{P}(\text{vaccinated control in week } j) &= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K \leq j) \\ \mathbb{P}(\text{unvaccinated control in week } j) &= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K > j). \end{aligned}$$

We can write the probability of being classified as a vaccinated control in week j for a given (x, u) as

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K \leq j | X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1 | K \leq j, X = x, U = u) \mathbb{P}(K \leq j | X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, K \leq j, X = x, U = u) \mathbb{P}(Y_j = 1 | K \leq j, X = x, U = u) \\
&\times \mathbb{P}(K \leq j | X = x, U = u),
\end{aligned}$$

and the probability of being classified as an unvaccinated control in week j for a given (x, u) as

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K > j | X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, K > j, X = x, U = u) \mathbb{P}(Y_j = 1 | K > j, X = x, U = u) \\
&\times \mathbb{P}(K > j | X = x, U = u).
\end{aligned}$$

If $Y_j = 1$, then the values of $Y_1, \ldots, Y_{j-1}$ can take on any value $y = 0, 1, 2$ since a person can have any number of NFARIs and may have had an FARI in a previous week. Therefore,

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, K \leq j, X = x, U = u) \\
&= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, \mathbf{Y_{j-1}} \leq \mathbf{2}, K \leq j, X = x, U = u) \\
&= \mathbb{P}(M_j = 1 | Y_j = 1, K \leq j, X = x, U = u) \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} \leq \mathbf{2}, K \leq j, X = x, U = u)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, K > j, X = x, U = u) \\
&= \mathbb{P}(M_j = 1 | Y_j = 1, K > j, X = x, U = u) \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} \leq \mathbf{2}, K > j, X = x, U = u)
\end{aligned}$$

where

$$\mathbb{P}(M_j = 1 | Y_j = 1, K \leq j, X = x, U = u) = \delta_{11u}$$

$$\mathbb{P}(M_j = 1 | Y_j = 1, K > j, X = x, U = u) = \delta_{10u}$$

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} \leq \mathbf{2}, K \leq j, X = x, U = u)$$
$$= \prod_{h=j_v}^{j-1} \mathbb{P}(M_h = 0 | Y_h \leq 2, K \leq j, X = x, U = u)$$

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} \leq \mathbf{2}, K > j, X = x, U = u)$$
$$= \prod_{h=j_v}^{j-1} \mathbb{P}(M_h = 0 | Y_h \leq 2, K > j, X = x, U = u).$$

Following a similar derivation as in section C.1.2, the probability of a vaccinated person not seeking care in week h ($M_h = 0$) is

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h \leq 2, K \leq j, X = x, U = u) \\
&= \frac{\mathbb{P}(M_h = 0, Y_h \leq 2, K \leq j | X = x, U = u)}{\mathbb{P}(Y_h \leq 2, K \leq j | X = x, U = u)} \\
&= \frac{\sum_{y=0}^{2} \sum_{g=j_v}^{j-1} \mathbb{P}(M_h = 0, Y_h = y, K = g | X = x, U = u)}{\sum_{y=0}^{2} \sum_{g=j_v}^{j-1} \mathbb{P}(Y_h = y, K = g | X = x, U = u)}
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(M_h = 0, Y_h = y, K = g | X = x, U = u) \\
&= \mathbb{P}(M_h = 0 | Y_h = y, K = g, X = x, U = u) \mathbb{P}(Y_h = y | K = g, X = x, U = u) \\
&\times \mathbb{P}(K = g | X = x, U = u)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h = 0, K = g, X = x, U = u) = 1 \\
&\mathbb{P}(M_h = 0 | Y_h = 1, K = g, X = x, U = u) = 1 - \delta_{11u} \qquad \text{when } g \leq h \\
&\mathbb{P}(M_h = 0 | Y_h = 1, K = g, X = x, U = u) = 1 - \delta_{10u} \qquad \text{when } g > h \\
&\mathbb{P}(M_h = 0 | Y_h = 2, K = g, X = x, U = u) = 1 - \delta_{21u} \qquad \text{when } g \leq h \\
&\mathbb{P}(M_h = 0 | Y_h = 2, K = g, X = x, U = u) = 1 - \delta_{20u} \qquad \text{when } g > h \\
&\mathbb{P}(Y_h = 0 | K = g, X = x, U = u) = 1 - \beta_{h1x} - \gamma_{h1x} \qquad \text{when } g \leq h \\
&\mathbb{P}(Y_h = 0 | K = g, X = x, U = u) = 1 - \beta_{h0x} - \gamma_{h0x} \qquad \text{when } g > h \\
&\mathbb{P}(Y_h = 1 | K = g, X = x, U = u) = \beta_{h1x} \qquad \text{when } g \leq h \\
&\mathbb{P}(Y_h = 1 | K = g, X = x, U = u) = \beta_{h0x} \qquad \text{when } g > h \\
&\mathbb{P}(Y_h = 2 | K = g, X = x, U = u) = \gamma_{h1x} \qquad \text{when } g \leq h \\
&\mathbb{P}(Y_h = 2 | K = g, X = x, U = u) = \gamma_{h0x} \qquad \text{when } g > h \\
&\mathbb{P}(K = g | X = x, U = u) = \alpha_{gxu}.
\end{aligned}$$

The probability of an unvaccinated person not seeking care in week h $(M_h = 0)$ is

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h \leq 2, K > j, X = x, U = u) \\
&= \frac{\mathbb{P}(M_h = 0, Y_h \leq 2, K > j | X = x, U = u)}{\mathbb{P}(Y_h \leq 2, K > j | X = x, U = u)} \\
&= \frac{\sum_{y=0}^{2} \mathbb{P}(M_h = 0, Y_h = y, K > j | X = x, U = u)}{\sum_{y=0}^{2} \mathbb{P}(Y_h = y, K > j | X = x, U = u)},
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(M_h = 0, Y_h = y, K > j | X = x, U = u) \\
&= \mathbb{P}(M_h = 0 | Y_h = y, K > j, X = x, U = u) \mathbb{P}(Y_h = y | K > j, X = x, U = u) \\
&\times \mathbb{P}(K > j | X = x, U = u)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h = 0, K > j, X = x, U = u) = 1 \\
&\mathbb{P}(M_h = 0 | Y_h = 1, K > j, X = x, U = u) = 1 - \delta_{10u} \\
&\mathbb{P}(M_h = 0 | Y_h = 2, K > j, X = x, U = u) = 1 - \delta_{20u} \\
&\mathbb{P}(Y_h = 0 | K > j, X = x, U = u) = 1 - \beta_{h0x} - \gamma_{h0x} \\
&\mathbb{P}(Y_h = 1 | K > j, X = x, U = u) = \beta_{h0x} \\
&\mathbb{P}(Y_h = 2 | K > j, X = x, U = u) = \gamma_{h0x} \\
&\mathbb{P}(K > j | X = x, U = u) = \sum_{g=j+1}^{J+1} \alpha_{gxu}.
\end{aligned}$$

The overall probability of being a vaccinated and an unvaccinated control, respectively, in week j are

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K \leq j)$$
$$= \sum_{x=0}^{1}\sum_{u=0}^{1} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K \leq j | X = x, U = u)\mathbb{P}(X = x, U = u)$$

or

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K > j)$$
$$= \sum_{x=0}^{1}\sum_{u=0}^{1} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K > j | X = x, U = u)\mathbb{P}(X = x, U = u).$$

For each week, we obtain the expected cell counts of the 2x2 table of cases and controls by vaccination status by multiplying N by the probability of being in each cell. For example the expected number of vaccinated cases in week j is $\mathbb{E}(\text{vaccinated cases in week } j) = N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K \leq j)$. Then, to obtain the final 2x2 table for a TN study, we sum the expected cell counts over all weeks during which vaccination occured $j_v, \ldots, J$. Finally, we can obtain the VE estimate from a TN study using the OR from the final 2x2 table. The VE estimate from a TN study is

$$VE_{TN} = 1 - OR_{TN} = \frac{\mathbb{E}(\text{vaccinated cases}) \times \mathbb{E}(\text{unvaccinated controls})}{\mathbb{E}(\text{unvaccinated cases}) \times \mathbb{E}(\text{vaccinated controls})},$$

where

$$\begin{aligned}
\mathbb{E}(\text{vaccinated cases}) &= \sum_{j=j_v}^{J} N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K \leq j) \\
\mathbb{E}(\text{unvaccinated cases}) &= \sum_{j=j_v}^{J} N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, K > j) \\
\mathbb{E}(\text{vaccinated controls}) &= \sum_{j=j_v}^{J} N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K \leq j) \\
\mathbb{E}(\text{unvaccinated controls}) &= \sum_{j=j_v}^{J} N \times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, K > j)
\end{aligned}$$

C.1.3 Input Parameters

Table C1 shows the parameter input values used in the calculations and simulations. Weekly values of the probabilities of FARI (γs) are based on weekly counts of positive influenza test results from the CDC during the 2009 influenza pandemic assuming an overall prevalence of 20% [7, 142]. Weekly counts of FARI after the introduction of a monovalent vaccine were adjusted assuming a VE of 62%. Weekly probabilities of NFARI (βs) were calculated based on the number of influenza negative specimens reported by the CDC during the 2009 pandemic. Weekly vaccination probabilities are based on vaccination coverage estimates from CDC vaccination surveillance during the 2009 pandemic [7].

Table C1: Input parameter values used in calculations and simulations.

Parameters	Input Values	Parameter	Input Values
π_{11}	0.4	α_{3711}	0.011
π_{10}	0.25	α_{3811}	0.0137
π_{01}	0.25	α_{3911}	0.0137
π_{11}	0.1	α_{4011}	0.0137
α_{111}	0.0	α_{4111}	0.00525
α_{211}	0.0	α_{4211}	0.00525
α_{311}	0.0	α_{4311}	0.00525
α_{411}	0.0	α_{4411}	0.00525
α_{511}	0.0	α_{4511}	0.002
α_{611}	0.0	α_{4611}	0.002
α_{711}	0.0	α_{4711}	0.002
α_{811}	0.0	α_{4811}	0.002
α_{911}	0.0	α_{4911}	0.002
α_{1011}	0.0	α_{5011}	0.001
α_{1111}	0.0	α_{5111}	0.001
α_{1211}	0.0	α_{5211}	0.001
α_{1311}	0.0	α_{5311}	0.001
α_{1411}	0.0	α_{5411}	0.00075
α_{1511}	0.0	α_{5511}	0.00075
α_{1611}	0.0	α_{5611}	0.00075
α_{1711}	0.0	α_{5711}	0.00075
α_{1811}	0.0	α_{110}	0.0
α_{1911}	0.0	α_{210}	0.0
α_{2011}	0.0	α_{310}	0.0
α_{2111}	0.0	α_{410}	0.0
α_{2211}	0.0	α_{510}	0.0
α_{2311}	0.0	α_{610}	0.0
α_{2411}	0.015	α_{710}	0.0
α_{2511}	0.015	α_{810}	0.0
α_{2611}	0.015	α_{910}	0.0
α_{2711}	0.015	α_{1010}	0.0
α_{2811}	0.0158	α_{1110}	0.0
α_{2911}	0.0158	α_{1210}	0.0
α_{3011}	0.0158	α_{1310}	0.0
α_{3111}	0.0158	α_{1410}	0.0
α_{3211}	0.0158	α_{1510}	0.0
α_{3311}	0.011	α_{1610}	0.0
α_{3411}	0.011	α_{1710}	0.0
α_{3511}	0.011	α_{1810}	0.0
α_{3611}	0.011	α_{1910}	0.0

Parameter	Input Values	Parameter	Input Values
α_{2010}	0.0	α_{101}	0.0
α_{2110}	0.0	α_{201}	0.0
α_{2210}	0.0	α_{301}	0.0
α_{2310}	0.0	α_{401}	0.0
α_{2410}	0.007	α_{501}	0.0
α_{2510}	0.007	α_{601}	0.0
α_{2610}	0.007	α_{701}	0.0
α_{2710}	0.007	α_{801}	0.0
α_{2810}	0.0088	α_{901}	0.0
α_{2910}	0.0088	α_{1001}	0.0
α_{3010}	0.0088	α_{1101}	0.0
α_{3110}	0.0088	α_{1201}	0.0
α_{3210}	0.0088	α_{1301}	0.0
α_{3310}	0.0072	α_{1401}	0.0
α_{3410}	0.0072	α_{1501}	0.0
α_{3510}	0.0072	α_{1601}	0.0
α_{3610}	0.0072	α_{1701}	0.0
α_{3710}	0.0072	α_{1801}	0.0
α_{3810}	0.01	α_{1901}	0.0
α_{3910}	0.01	α_{2001}	0.0
α_{4010}	0.01	α_{2101}	0.0
α_{4110}	0.00425	α_{2201}	0.0
α_{4210}	0.00425	α_{2301}	0.0
α_{4310}	0.00425	α_{2401}	0.021
α_{4410}	0.00425	α_{2501}	0.021
α_{4510}	0.0014	α_{2601}	0.021
α_{4610}	0.0014	α_{2701}	0.021
α_{4710}	0.0014	α_{2801}	0.0216
α_{4810}	0.0014	α_{2901}	0.0216
α_{4910}	0.0014	α_{3001}	0.0216
α_{5010}	0.00075	α_{3101}	0.0216
α_{5110}	0.00075	α_{3201}	0.0216
α_{5210}	0.00075	α_{3301}	0.0128
α_{5310}	0.00075	α_{3401}	0.0128
α_{5410}	0.0005	α_{3501}	0.0128
α_{5510}	0.0005	α_{3601}	0.0128
α_{5610}	0.0005	α_{3701}	0.0128
α_{5710}	0.0005		

Parameter	Input Values	Parameter	Input Values
α_{3801}	0.015	α_{2100}	0.0
α_{3901}	0.015	α_{2200}	0.0
α_{4001}	0.015	α_{2300}	0.0
α_{4101}	0.00575	α_{2400}	0.01425
α_{4201}	0.00575	α_{2500}	0.01425
α_{4301}	0.00575	α_{2600}	0.01425
α_{4401}	0.00575	α_{2700}	0.01425
α_{4501}	0.0018	α_{2800}	0.0134
α_{4601}	0.0018	α_{2900}	0.0134
α_{4701}	0.0018	α_{3000}	0.0134
α_{4801}	0.0018	α_{3100}	0.0134
α_{4901}	0.0018	α_{3200}	0.0134
α_{5001}	0.00125	α_{3300}	0.0108
α_{5101}	0.00125	α_{3400}	0.0108
α_{5201}	0.00125	α_{3500}	0.0108
α_{5301}	0.00125	α_{3600}	0.0108
α_{5401}	0.001	α_{3700}	0.0108
α_{5501}	0.001	α_{3800}	0.0197
α_{5601}	0.001	α_{3900}	0.0197
α_{5701}	0.001	α_{4000}	0.0197
α_{100}	0.0	α_{4100}	0.007
α_{200}	0.0	α_{4200}	0.007
α_{300}	0.0	α_{4300}	0.007
α_{400}	0.0	α_{4400}	0.007
α_{500}	0.0	α_{4500}	0.0026
α_{600}	0.0	α_{4600}	0.0026
α_{700}	0.0	α_{4700}	0.0026
α_{800}	0.0	α_{4800}	0.0026
α_{900}	0.0	α_{4900}	0.0026
α_{1000}	0.0	α_{5000}	0.00175
α_{1100}	0.0	α_{5100}	0.00175
α_{1200}	0.0	α_{5200}	0.00175
α_{1300}	0.0	α_{5300}	0.00175
α_{1400}	0.0	α_{5400}	0.00075
α_{1500}	0.0	α_{5500}	0.00075
α_{1600}	0.0	α_{5600}	0.00075
α_{1700}	0.0	α_{5700}	0.00075
α_{1800}	0.0	β_{101}	0000256009
α_{1900}	0.0	β_{201}	0.002377228
α_{2000}	0.0	β_{301}	0.004374099

Parameter	Input Values	Parameter	Input Values
β_{401}	0.003103197	β_{4401}	0.000476993
β_{501}	0.004754456	β_{4501}	0.000578566
β_{601}	0.005138469	β_{4601}	0.000618846
β_{701}	0.006634294	β_{4701}	0.000657295
β_{801}	0.007137169	β_{4801}	0.000452233
β_{901}	0.007989314	β_{4901}	0.000314916
β_{1001}	0.005719976	β_{5001}	0.000287274
β_{1101}	0.004313754	β_{5101}	0.000171998
β_{1201}	0.004706911	β_{5201}	0.000100637
β_{1301}	0.003783449	β_{5301}	0.0000292763
β_{1401}	0.003348234	β_{5401}	0.0000237833
β_{1501}	0.003386635	β_{5501}	0.0000329308
β_{1601}	0.002327855	β_{5601}	0.00000914744
β_{1701}	0.002285796	β_{5701}	0.00000365897
β_{1801}	0.002885589	θ_{β}	0.5-2.0
β_{1901}	0.002479631	ϕ_{β}	1.0-2.0
β_{2001}	0.003445152	γ_{101}	0.007128339
β_{2101}	0.004194893	γ_{201}	0.056857479
β_{2201}	0.004785542	γ_{301}	0.052542095
β_{2301}	0.005228073	γ_{401}	0.024699552
β_{2401}	0.007100189	γ_{501}	0.020408465
β_{2501}	0.011297181	γ_{601}	0.016159923
β_{2601}	0.015924209	γ_{701}	0.014700962
β_{2701}	0.017967336	γ_{801}	0.014375775
β_{2801}	0.014288605	γ_{901}	0.014491285
β_{2901}	0.011232564	γ_{1001}	0.014058863
β_{3001}	0.00854768	γ_{1101}	0.013251482
β_{3101}	0.005013384	γ_{1201}	0.012829935
β_{3201}	0.002599943	γ_{1301}	0.01257947
β_{3301}	0.001835585	γ_{1401}	0.011016051
β_{3401}	0.001123076	γ_{1501}	0.011766687
β_{3501}	0.000883732	γ_{1601}	0.011122297
β_{3601}	0.000462118	γ_{1701}	0.011465152
β_{3701}	0.000524716	γ_{1801}	0.012859199
β_{3801}	0.000490549	γ_{1901}	0.01607784
β_{3901}	0.000481328	γ_{2001}	0.020914872
β_{4001}	0.000584602	γ_{2101}	0.023880692
β_{4101}	0.00049167	γ_{2201}	0.026345791
β_{4201}	0.000532031	γ_{2301}	0.027179623
β_{4301}	0.000436632	γ_{2401}	0.027319149

Parameter	Input Values	Parameter	Input Values
γ_{2501}	0.030194027	γ_{4701}	0.012814887
γ_{2601}	0.03351409	γ_{4801}	0.010986937
γ_{2701}	0.036411291	γ_{4901}	0.009645197
γ_{2801}	0.033578504	γ_{5001}	0.008982453
γ_{2901}	0.031866133	γ_{5101}	0.007833009
γ_{3001}	0.029362385	γ_{5201}	0.006903403
γ_{3101}	0.027185619	γ_{5301}	0.006556295
γ_{3201}	0.020909266	γ_{5401}	0.006324788
γ_{3301}	0.021578301	γ_{5501}	0.005014685
γ_{3401}	0.019280229	γ_{5601}	0.003975537
γ_{3501}	0.018226878	γ_{5701}	0.002058121
γ_{3601}	0.014984079	θ_γ	0.34
γ_{3701}	0.016840602	ϕ_γ	1.0-2.0
γ_{3801}	0.016882373	δ_{101}	0.25
γ_{3901}	0.015156862	θ_{δ_1}	1.0
γ_{4001}	0.014944917	μ_{δ_1}	0.5-1.0
γ_{4101}	0.014932952	δ_{201}	0.40
γ_{4201}	0.014698786	θ_{δ_2}	0.5-1.0
γ_{4301}	0.014390479	μ_{δ_2}	0.5-1.0
γ_{4401}	0.014707195	τ_1	0
γ_{4501}	0.013856789	τ_2	1
γ_{4601}	0.013078778		

We assume $\mu_{\delta_1} = \mu_{\delta_2}$.

C.1.4 Results

We evaluated bias of VE estimates from TN studies in the presence of different sources of bias (Table 4.4). Table C2 shows the 5^{th}, 50^{th}, and 95^{th} quantiles of bias and the 50^{th} and 95^{th} quantiles of the absolute value of bias under combination of sources of bias.

Table C2: Bias of TN-based estimates of VE when first week of vaccination varies: multiple sources of bias

Source of Bias	**Outcome of Interest**	j_v	$\mathbf{Q}_B$**(5)**	$\mathbf{Q}_B$**(50)**	$\mathbf{Q}_B$**(95)**	$\mathbf{Q}_{AVB}$**(50)**	$\mathbf{Q}_{AVB}$**(95)**
BS, D	SI & MAI	1 [c]	0.00	0.00	0.00	0.00	0.00
		6	0.05	0.05	0.05	0.05	0.05
		11	0.10	0.10	0.10	0.10	0.10
		15	0.05	0.05	0.05	0.05	0.05
		20	-0.03	-0.03	-0.03	0.03	0.03
		24	-0.04	-0.04	-0.04	0.04	0.04
		28	-0.03	-0.03	-0.03	0.03	0.03
		33	-0.05	-0.05	-0.05	0.05	0.05
		38	-0.08	-0.08	-0.07	0.08	0.08
BS, D, A	SI & MAI	1	-0.22	0.01	0.15	0.08	0.22
		6	-0.15	0.04	0.17	0.07	0.18
		11	-0.16	0.09	0.25	0.11	0.25
		15	-0.17	0.05	0.19	0.09	0.21
		20	-0.22	-0.03	0.09	0.06	0.22
		24	-0.17	-0.04	0.04	0.05	0.17
		28	-0.13	-0.03	0.03	0.03	0.13
		33	-0.25	-0.06	0.08	0.07	0.25
		38	-0.28	-0.08	0.05	0.08	0.28
BS, D, C	SI	1	0.01	0.08	0.16	0.08	0.16
		6	0.06	0.11	0.19	0.11	0.19
		11	0.10	0.17	0.27	0.17	0.27
		15	0.05	0.11	0.21	0.11	0.21
		20	-0.03	0.03	0.11	0.03	0.11
		24	-0.04	0.00	0.05	0.02	0.05
		28	-0.03	0.00	0.04	0.02	0.04
		33	-0.04	0.01	0.09	0.03	0.09
		38	-0.07	-0.01	0.06	0.04	0.07
	MAI	1	0.00	0.00	0.00	0.00	0.00
		6	0.03	0.04	0.05	0.04	0.05
		11	0.06	0.08	0.09	0.08	0.09
		15	0.03	0.04	0.05	0.04	0.05
		20	-0.03	-0.02	-0.02	0.02	0.03
		24	-0.04	-0.03	-0.02	0.03	0.04
		28	-0.03	-0.02	-0.02	0.02	0.03
		33	-0.04	-0.04	-0.03	0.04	0.04
		38	-0.07	-0.06	-0.04	0.06	0.07

Table C2: Bias of TN-based estimates of VE when first week of vaccination varies: multiple sources of bias

Source of Bias	**Outcome of Interest**	j_v	$\mathbf{Q_B(5)}$	$\mathbf{Q_B(50)}$	$\mathbf{Q_B(95)}$	$\mathbf{Q_{AVB}(50)}$	$\mathbf{Q_{AVB}(95)}$
BS, D, A, C	SI	1	-0.13	0.08	0.21	0.10	0.2
		6	-0.08	0.12	0.22	0.12	0.22
		11	-0.04	0.18	0.31	0.18	0.31
		15	-0.10	0.12	0.25	0.13	0.26
		20	-0.16	0.03	0.15	0.07	0.17
		24	-0.11	0.01	0.08	0.04	0.11
		28	-0.09	0.00	0.06	0.03	0.09
		33	-0.16	0.03	0.13	0.07	0.17
		38	-0.20	-0.01	0.11	0.06	0.20
	MAI	1	-0.18	0.00	0.12	0.06	0.18
		6	-0.13	0.04	0.15	0.06	0.16
		11	-0.11	0.07	0.21	0.09	0.21
		15	-0.15	0.04	0.16	0.07	0.18
		20	-0.19	-0.03	0.07	0.05	0.19
		24	-0.14	-0.03	0.03	0.03	0.14
		28	-0.11	-0.02	0.03	0.03	0.11
		33	-0.20	-0.03	0.07	0.05	0.20
		38	-0.23	-0.06	0.04	0.06	0.23

C.2 All-Or-None Vaccine Model

C.2.1 True VE

True VE against SI

A person is considered a true case of SI if s/he develops an ARI as a result of influenza infection ($Y_j = 2$). True VE against SI is defined as one minus the risk of influenza infection among vaccinated persons compared to unvaccinated persons.

$$VE_{TSI} = 1 - \frac{\mathbb{P}(\text{classified as SI}|V = 1)}{\mathbb{P}(\text{classified as SI}|V = 0)}$$

We can write probability of influenza infection in week j for a given (v, x, u) as:

$$\mathbb{P}(Y_j = 2|V = v, X = x, U = u)$$
$$= \sum_{w=0}^{1} \mathbb{P}(Y_j = 2|V = v, W = w, X = x, U = u)\mathbb{P}(W = w|V = v)$$

where

$$\mathbb{P}(Y_j = 2|V = 0, W = w, X = x, U = u) = \gamma_{j0x}$$
$$\mathbb{P}(Y_j = 2|V = 1, W = 0, X = x, U = u) = \gamma_{j0x}$$
$$\mathbb{P}(Y_j = 2|V = 1, W = 1, X = x, U = u) = 0$$
$$\mathbb{P}(V = 0|X = x, U = u) = 1 - \alpha_{xu}$$
$$\mathbb{P}(V = 1|X = x, U = u) = \alpha_{xu}$$
$$\mathbb{P}(W = 0|V = 0) = 1$$
$$\mathbb{P}(W = 1|V = 0) = 0$$
$$\mathbb{P}(W = 0|V = 1) = 1 - \rho$$
$$\mathbb{P}(W = 1|V = 1) = \rho.$$

The probability of influenza infection in week j can be expressed as

$$\begin{aligned}&\mathbb{P}(Y_j = 2|V = v)\\ &= \sum_{x=0}^{1}\sum_{u=0}^{1}\mathbb{P}(Y_j = 2|W = 0, V = v, X = x, U = u)\mathbb{P}(W = 0|V = v)\mathbb{P}(X = x, U = u|V = v)\end{aligned}$$

where

$$\mathbb{P}(X = x, U = u|V = 0) = \frac{(1-\alpha_{xu})\pi_{xu}}{\sum_{x=0}^{1}\sum_{u=0}^{1}(1-\alpha_{xu})\pi_{xu}}$$

$$\mathbb{P}(X = x, U = u|V = 1) = \frac{\alpha_{xu}\pi_{xu}}{\sum_{x=0}^{1}\sum_{u=0}^{1}\alpha_{xu}\pi_{xu}}$$

Under the assumption of random vaccination ($\alpha_{xu} = \alpha \forall x, u$ ($x = 0, 1; u = 0, 1$)), the probability of a given (x, u) given vaccination status v can be further reduced to

$$\mathbb{P}(X = x, U = u|V = 0) = \frac{(1-\alpha)\pi_{xu}}{\sum_{x=0}^{1}\sum_{u=0}^{1}(1-\alpha)\pi_{xu}} = \pi_{xu}$$

$$\mathbb{P}(X = x, U = u|V = 1) = \frac{\alpha\pi_{xu}}{\sum_{x=0}^{1}\sum_{u=0}^{1}\alpha\pi_{xu}} = \pi_{xu}$$

Therefore, the true VE against SI is

$$\begin{aligned}VE_{TSI} &= 1 - \frac{\sum_{j=1}^{J}(1-\rho)\left[\gamma_{j00}(\pi_{00}+\pi_{01}) + \gamma_{j01}(\pi_{10}+\pi_{11})\right]}{\sum_{j=1}^{J}\gamma_{j00}\left[\pi_{00}+\pi_{01}\right] + \gamma_{j01}\left[\pi_{10}+\pi_{11}\right]}\\ &= 1-(1-\rho)\\ &= \rho\end{aligned}$$

True VE against MAI

A person is considered a true case of MAI in week j if s/he develops an ARI as a result of influenza infection and seeks medical care for this ARI ($Y_j = 2, M_j = 1$). We can write probability of seeking medical care for influenza infection in week j for a given (v, x, u) as:

$$\mathbb{P}(M_j = 1, Y_j = 2|V = v, X = x, U = u)$$
$$= \sum_{w=0}^{1} \mathbb{P}(M_j = 1, Y_j = 2|W = w, V = v, X = x, U = u)\mathbb{P}(W = w|V = v)$$

where

$$\mathbb{P}(M_j = 1, Y_j = 2|V = v, W = w, X = x, U = u)$$
$$= \mathbb{P}(M_j = 1|Y_j = 2, W = w, V = v, X = x, U = u)\mathbb{P}(Y_j = 2|W = w, V = v, X = x, U = u)$$

and

$$\mathbb{P}(M_j = 1|Y_j = 2, W = 0, V = v, X = x, U = u) = \delta_{20u}$$
$$\mathbb{P}(M_j = 1|Y_j = 2, W = 1, V = v, X = x, U = u) = 0$$
$$\mathbb{P}(Y_j = 2|W = 0, V = v, X = x, U = u) = \gamma_{j0x}$$
$$\mathbb{P}(Y_j = 2|W = 1, V = v, X = x, U = u) = 0$$
$$\mathbb{P}(W = 0|V = 0) = 1$$
$$\mathbb{P}(W = 1|V = 0) = 0$$
$$\mathbb{P}(W = 0|V = 1) = 1 - \rho$$
$$\mathbb{P}(W = 1|V = 1) = \rho.$$

The probability of seeking medical care for influenza infection in week j can be expressed as

$$\begin{aligned}&\mathbb{P}(M_j = 1, Y_j = 2|V = v)\\&= \sum_{x=0}^{1}\sum_{u=0}^{1}\mathbb{P}(M_j = 1, Y_j = 2|W = 0, V = v, X = x, U = u)\mathbb{P}(W = 0|V = v)\mathbb{P}(X = x, U = u|V = v)\end{aligned}$$

where

$\mathbb{P}(X = x, U = u|V = v) = \pi_{xu},$ assuming random vaccination.

Therefore, the true VE against MAI is

$$\begin{aligned}VE_{TMAI} &= 1 - \frac{\sum_{j=1}^{J}(1-\rho)\left[\delta_{210}(\gamma_{j00}\pi_{00} + \gamma_{j01}\pi_{10}) + \delta_{211}(\gamma_{j00}\pi_{01} + \pi_{11})\right]}{\sum_{j=1}^{J}\delta_{200}\left[\gamma_{j00}\pi_{00} + \gamma_{j01}\pi_{10}\right] + \delta_{201}\left[\gamma_{j00}\pi_{01} + \gamma_{j01}\pi_{11}\right]}\\&= 1 - (1-\rho)\\&= \rho\end{aligned}$$

C.2.2 TN-based Estimates of VE

Case

The probability of being a case in week j is the probability of having FARI in week j ($Y_j = 2$), seeking medical care for FARI in week j ($M_j = 1$), and not seeking medical care prior to week j (i.e., a person's first medical visit occurs in week j, $\mathbf{M_{j-1}} = \mathbf{0}$). We assume perfect influenza test sensitivity and specificity. We can write the probability of being a case in week j as

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, V = v), \quad v = 0, 1$$

For a given (x, u), we can write the probability of being a case in week j as

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, V = v | X = x, U = u)$$
$$= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2 | V = v, X = x, U = u)\mathbb{P}(V = v | X = x, U = u)$$

where

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2 | V = v, X = x, U = u)$$
$$= \sum_{w=0}^{1} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2 | W = w, V = v, X = x, U = u)\mathbb{P}(W = w | V = v)$$

and

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2 | W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, W = w, V = v, X = x, U = u)$$
$$\times \mathbb{P}(Y_j = 2 | V = v, W = w, X = x, U = u)$$

Under the assumption that a person can have only one FARI during the study,

$$\mathbb{P}(Y_j = 2 | V = v, W = w, X = x, U = u)$$
$$= \mathbb{P}(\mathbf{Y_{j-1}} < \mathbf{2}, Y_j = 2 | V = v, W = w, X = x, U = u)$$
$$= \gamma_{jwx}.$$

Therefore,

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 2, W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | \mathbf{Y_{j-1}} < \mathbf{2}, Y_j = 2, W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(M_j = 1 | Y_j = 2, W = w, V = v, X = x, U = u)$$
$$\times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, W = w, V = v, X = x, U = u),$$

where

$$\mathbb{P}(M_j = 1 | Y_j = 2, W = 0, V = v, X = x, U = u) = \delta_{20u}$$
$$\mathbb{P}(M_j = 1 | Y_j = 2, W = 1, V = v, X = x, U = u) = 0$$
$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} < \mathbf{2}, W = w, V = v, X = x, U = u)$$
$$= \prod_{h=1}^{j-1} \mathbb{P}(M_h = 0 | Y_h < 2, W = w, V = v, X = x, U = u),$$

since $M_1, \ldots, M_{j-1}$ are independent given $Y_1, \ldots, Y_{j-1}$.

The probability of not seeking care in week h ($M_h = 0$) given no FARI ($Y_h < 2$) can be derived as follows:

$$\begin{aligned}
&\mathbb{P}(M_h = 0 | Y_h < 2, W = w, V = v, X = x, U = u) \\
&= \frac{\mathbb{P}(M_h = 0, Y_h < 2, W = w, V = v, X = x, U = u)}{\mathbb{P}(Y_h < 2, W = w, V = v, X = x, U = u)} \\
&= \frac{\sum_{y=0}^{1} \mathbb{P}(M_h = 0, Y_h = y, W = w, V = v, X = x, U = u)}{\sum_{y=0}^{1} \mathbb{P}(Y_h = y, W = w, V = v, X = x, U = u)}
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(M_h = 0, Y_h = y, W = w, V = v, X = x, U = u) \\
&= \mathbb{P}(M_h = 0 | Y_h = y, W = w, V = v, X = x, U = u) \mathbb{P}(Y_h = y | W = w, V = v, X = x, U = u) \\
&\times \mathbb{P}(W = w | V = v) \mathbb{P}(V = v | X = x, U = u) \mathbb{P}(X = x, U = u)
\end{aligned}$$

and

$$\mathbb{P}(M_h = 0 | Y_h = 0, W = w, V = v, X = x, U = u) = 1$$

$$\mathbb{P}(M_h = 0 | Y_h = 1, W = 0, V = v, X = x, U = u) = 1 - \delta_{10u}$$

$$\mathbb{P}(M_h = 0 | Y_h = 1, W = 1, V = 0, X = x, U = u) = 0$$

$$\mathbb{P}(M_h = 0 | Y_h = 1, W = 1, V = 1, X = x, U = u) = 1 - \delta_{11u}$$

$$\mathbb{P}(Y_h = 0 | W = 0, V = v, X = x, U = u) = 1 - \beta_{h0x} - \gamma_{h0x}$$

$$\mathbb{P}(Y_h = 0 | W = 1, V = 1, X = x, U = u) = 1 - \beta_{h1x}$$

$$\mathbb{P}(Y_h = 1 | W = 0, V = v, X = x, U = u) = \beta_{h0x}$$

$$\mathbb{P}(Y_h = 1 | W = 1, V = 1, X = x, U = u) = \beta_{h1x}$$

$$\mathbb{P}(W = 0 | V = 0) = 1$$

$$\mathbb{P}(W = 1 | V = 0) = 0$$

$$\mathbb{P}(W = 0 | V = 1) = 1 - \rho$$

$$\mathbb{P}(W = 1 | V = 1) = \rho$$

$$\mathbb{P}(V = 0 | X = x, U = u) = 1 - \alpha_{xu}$$

$$\mathbb{P}(V = 1 | X = x, U = u) = \alpha_{xu}$$

$$\mathbb{P}(X = x, U = u) = \pi_{xu}$$

Therefore,

$$\mathbb{P}(M_h = 0 | Y_h < 2, W = 0, V = v, X = x, U = u) = \frac{1 - \delta_{10u}\beta_{h0x} - \gamma_{h0x}}{1 - \gamma_{h0x}}$$

The overall probability of being a case in week j is

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, V = 0) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1} \delta_{20u}\gamma_{j0x}(1 - \alpha_{xu})\pi_{xu} \prod_{h=1}^{j-1} \frac{1 - \delta_{10u}\beta_{h0x} - \gamma_{h0x}}{1 - \gamma_{h0x}}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 2, V = 1) \\
&= \sum_{x=0}^{1}\sum_{u=0}^{1} \delta_{20u}\gamma_{j0x}(1 - \rho)\alpha_{xu}\pi_{xu} \prod_{h=1}^{j-1} \frac{1 - \delta_{10u}\beta_{h0x} - \gamma_{h0x}}{1 - \gamma_{h0x}}
\end{aligned}$$

Control

A person is classified as a control in week j in the same way as a case, except that they seek medical care for NFARI ($\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1$). We can write the probability of being a control in week j as

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, V = v), \quad v = 0, 1$$

For a given (x, u), we can write the probability of being a control in week j as

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, V = v | X = x, U = u)$$
$$= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1 | V = v, X = x, U = u)\mathbb{P}(V = v | X = x, U = u)$$

where

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1 | V = v, X = x, U = u)$$
$$= \sum_{w=0}^{1} \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1 | W = w, V = v, X = x, U = u)\mathbb{P}(W = w | V = v)$$

and

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1 | W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, W = w, V = v, X = x, U = u)$$
$$\times \mathbb{P}(Y_j = 1 | V = v, W = w, X = x, U = u)$$

If $Y_j = 1$, then the values of $Y_1, \ldots, Y_{j-1}$ can take on any value $y = 0, 1, 2$ since a person can have any number of NFARIs and may have had an FARI in a previous week. Therefore,

$$\mathbb{P}(Y_j = 1 | V = v, W = w, X = x, U = u)$$
$$= \mathbb{P}(\mathbf{Y_{j-1}} \leq \mathbf{2}, Y_j = 1 | V = v, W = w, X = x, U = u)$$

Thus,

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | Y_j = 1, W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1 | \mathbf{Y_{j-1}} \leq \mathbf{2}, Y_j = 1, W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(M_j = 1 | Y_j = 1, W = w, V = v, X = x, U = u)$$
$$\times \mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} \leq \mathbf{2}, W = w, V = v, X = x, U = u),$$

where

$$\mathbb{P}(M_j = 1 | Y_j = 1, W = 0, V = v, X = x, U = u) = \delta_{10u}$$
$$\mathbb{P}(M_j = 1 | Y_j = 1, W = 1, V = 1, X = x, U = u) = \delta_{11u}$$
$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0} | \mathbf{Y_{j-1}} \leq \mathbf{2}, W = w, V = v, X = x, U = u)$$
$$= \prod_{h=1}^{j-1} \mathbb{P}(M_h = 0 | Y_h \leq 2, W = w, V = v, X = x, U = u),$$

The probability of not seeking care in week h ($M_h = 0$) can be derived as follows:

$$\mathbb{P}(M_h = 0|Y_h \leq 2, W = w, V = v, X = x, U = u)$$
$$= \frac{\mathbb{P}(M_h = 0, Y_h \leq 2, W = w, V = v, X = x, U = u)}{\mathbb{P}(Y_h \leq 2, W = w, V = v, X = x, U = u)}$$
$$= \frac{\sum_{y=0}^{2} \mathbb{P}(M_h = 0, Y_h = y, W = w, V = v, X = x, U = u)}{\sum_{y=0}^{2} \mathbb{P}(Y_h = y, W = w, V = v, X = x, U = u)}$$

where

$$\mathbb{P}(M_h = 0, Y_h = y, W = w, V = v, X = x, U = u)$$
$$= \mathbb{P}(M_h = 0|Y_h = y, W = w, V = v, X = x, U = u)\mathbb{P}(Y_h = y|W = w, V = v, X = x, U = u)$$
$$\times \mathbb{P}(W = w|V = v)\mathbb{P}(V = v|X = x, U = u)\mathbb{P}(X = x, U = u)$$

and

$$\mathbb{P}(M_h = 0|Y_h = 0, W = w, V = v, X = x, U = u) = 1$$
$$\mathbb{P}(M_h = 0|Y_h = 1, W = 0, V = v, X = x, U = u) = 1 - \delta_{10u}$$
$$\mathbb{P}(M_h = 0|Y_h = 1, W = 1, V = 0, X = x, U = u) = 0$$
$$\mathbb{P}(M_h = 0|Y_h = 1, W = 1, V = 1, X = x, U = u) = 1 - \delta_{11u}$$
$$\mathbb{P}(M_h = 0|Y_h = 2, W = 0, V = v, X = x, U = u) = 1 - \delta_{20u}$$
$$\mathbb{P}(M_h = 0|Y_h = 2, W = 1, V = v, X = x, U = u) = 0$$
$$\mathbb{P}(Y_h = 0|W = 0, V = v, X = x, U = u) = 1 - \beta_{h0x} - \gamma_{h0x}$$
$$\mathbb{P}(Y_h = 0|W = 1, V = 1, X = x, U = u) = 1 - \beta_{h1x}$$
$$\mathbb{P}(Y_h = 1|W = 0, V = v, X = x, U = u) = \beta_{h0x}$$
$$\mathbb{P}(Y_h = 1|W = 1, V = 1, X = x, U = u) = \beta_{h1x}$$
$$\mathbb{P}(Y_h = 2|W = 0, V = v, X = x, U = u) = \gamma_{h0x}$$
$$\mathbb{P}(Y_h = 2|W = 1, V = 1, X = x, U = u) = 0$$
$$\mathbb{P}(W = 0|V = 0) = 1$$
$$\mathbb{P}(W = 1|V = 0) = 0$$

$$\mathbb{P}(W = 0|V = 1) = 1 - \rho$$
$$\mathbb{P}(W = 1|V = 1) = \rho$$
$$\mathbb{P}(V = 0|X = x, U = u) = 1 - \alpha_{xu}$$
$$\mathbb{P}(V = 1|X = x, U = u) = \alpha_{xu}$$
$$\mathbb{P}(X = x, U = u) = \pi_{xu}$$

The overall probability of being a control in week j is

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, V = 0)$$
$$= \sum_{x=0}^{1}\sum_{u=0}^{1} \delta_{20u}\beta_{j0x}(1 - \alpha_{xu})\pi_{xu} \prod_{h=1}^{j-1}(1 - \delta_{10u}\beta_{h0x} - \delta_{20u})$$

and

$$\mathbb{P}(\mathbf{M_{j-1}} = \mathbf{0}, M_j = 1, Y_j = 1, V = 1)$$
$$= \sum_{x=0}^{1}\sum_{u=0}^{1} \left\{ \alpha_{xu}\pi_{xu} \left[\delta_{10u}\beta_{j0x}(1 - \rho) \prod_{h=1}^{j-1}(1 - \delta_{10u}\beta_{h0x} - \delta_{20u}) + \delta_{11u}\beta_{j1x}\rho \prod_{h=1}^{j-1}(1 - \delta_{11u}\beta_{h1x}) \right] \right\}$$

References

1. World Health Organization. *Influenza (Seasonal)* http://www.who.int/mediacentre/factsheets/fs211/en/ (2018).
2. Barker, W. & Mullooly, J. Impact of epidemic type A influenza in a defined adult population. *American Journal of Epidemiology* **112,** 798–811 (1980).
3. Barker, W. Excess pneumonia and influenza associated hospitalization during influenza epidemics in the United States, 1970-78. *American Journal of Public Health* **76,** 761–765 (1986).
4. Poehling, K. *et al.* The underrecognized burden of influenza in young children. *NEJM* **355,** 31–40 (2006).
5. Centers for Disease Control and Prevention. Estimates of Deaths Associated with Seasonal Influenza – United States, 1976-2007. *Morbidity and Mortality Weekly Report* **59,** 1057–1062 (2010).
6. Paules, C. I., Sullivan, S. G., Subbarao, K. & Fauci, A. S. Chasing seasonal influenza - the need for a universal influenza vaccine. *New England Journal of Medicine* (2017).
7. Centers for Disease Control and Prevention. *Influenza (flu)* http://www.cdc.gov/flu/index.htm (2017).
8. Centers for Disease Control and Prevention. *Estimated Influenza Illnesses and Hospitalizations Averted by Vaccination — United States, 2014–15 Influenza Season* http://www.cdc.gov/flu/about/disease/2014-15.htm (2016).
9. Srestha, S. *et al.* Estimating the burden of 2009 pandemic influenza A (H1N1) in the United States (April 2009-April 2010). *Clin Infect Dis* **52 Suppl 1,** S75–82 (2011).
10. Borse, R. *et al.* Effects of Vaccine Program against Pandemic Influenza A(H1N1) Virus, United States, 2009–2010. *Emerging Infectious Diseases* **19,** 439–448 (3 2013).
11. Centers for Disease Control and Prevention. Prevention and control of influenza with vaccines: recommendations from the Advisory Committee on Immunization Practices (ACIP). *Morbidity and Mortality Weekly Report* **59** (2010).

12. Rolfes, M. *et al. Estimated Influenza Illnesses, Medical Visits, Hospitalizations, and Deaths Averted by Vaccination in the United States* https://www.cdc.gov/flu/about/disease/2015-16.htm (2018).

13. Palese, P. & Shah, M. in (eds Knipe, D. & Howley, P.) 5th ed., 1647–1690 (Lippincott, Williams and Wilkins, Philadelphia, 2007).

14. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens* **3,** 1470–1476 (2007).

15. Taubenberger, J. K. & Morens, D. M. The pathology of influenza virus infections. *Annual Reviews of Pathology* **3,** 499–522 (2008).

16. Obenauer, J. *et al.* Large-scale sequence analysis of avian influenza isolates. *Science* **311,** 1576–1580 (2006).

17. Webster, R., Bean, W., Gorman, O., Chambers, T. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol Rev* **56,** 152–179 (1992).

18. Centers for Disease Control and Prevention. Update: influenza activity - United States and Worldwide, May 20 - September 17, 2007. *Morbidity and Mortality Weekly Report* **55,** 1001–1004 (2007).

19. Tellier, R. Review of aerosol transmission of influenza A virus. *Emerging Infectious Diseases* **12,** 1657–1662 (2006).

20. Shek, L. P. & Lee, B. W. Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric Respiratory Reviews* **4,** 105–111 (2003).

21. Viboud, C., Alonso, W. J. & Simonsen, L. Influenza in tropical regions. *PLoS Medicine* **3,** e89. doi:10.1371/journal.pmed.0030089 (2006).

22. Tamerius, J. D. *et al.* Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathogens* **9,** e1003194. doi:10.1371/journal.ppat.1003194 (2013).

23. Lowen, A. C., Steel, J., Mubareka, S. & Palese, P. High temperature (30°C) blocks aerosol but not contact transmission of influenza virus. *Journal of Virology* **78,** 5650–5652 (2008).

24. Cowling, B. J. *et al.* Aerosol transmission is an important mode of influenza A virus spread. *Nature Communications* **4,** doi:10.1038/ncomms2922 (2013).

25. Hay, H. J., Gregory, V., Douglas, A. R. & Lin, Y. P. The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci* **356,** 1861–1870 (2001).

26. Centers for Disease Control and Prevention. *How the Flu Virus Can Change: "Drift" and "Shift"* http://www.cdc.gov/flu/about/viruses/change.htm (2016).

27. Johnson, N. P. & Mueller, J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza. *Bulletin of the History of Medicine* **76,** 105–115 (1 2002).

28. Morens, D. M. & Fauci, A. S. The 1918 influenza pandemic: insights for the 21st century. *Journal of Infectious Diseases* **195,** 1018–1028 (2007).

29. Jhung, M. A. *et al.* Epidemiology of 2009 pandemic influenza A (H1N1) in the United States. *Clinical Infectious Diseases* **52,** S13–S26 (Suppl 1 2011).

30. Yen, H.-L. & Webster, R. G. in (eds Compans, R. W. & Orenstein, W. A.) chap. Pandemic Influenza as a Current Threat (Springer Berlin Heidelberg, 2009).

31. Centers for Disease Control and Prevention. *Key Facts About Seasonal Flu Vaccine* http://www.cdc.gov/flu/protect/keyfacts.htm (2016).

32. Foppa, I. M. *et al.* Deaths averted by influenza vaccination in the U.S. during the seasons 2005/06 through 2013/14. *Vaccine* **33,** 3003–3009 (26 2015).

33. Hannoun, C. The evolving history of influenza viruses and influenza vaccines. *Expert Rev Vaccines* **12,** 1085–1094 (2013).

34. Meiklejon, G. in. Chap. Commission on influenza (The Borden Institute, Office of the Surgeon General, Department fo the Army, Falls Church, VA, 1994).

35. Osterholm, M. T., Kelley, N. S., Sommer, A. & Belongia, E. A. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infectious Diseases* **12,** 36–44 (2012).

36. Panthier, R., Cateigne, G. & Hannoun, C. Isolation of influenza virus strain. Reaction of a young monkey following intra-nasal innoculation with this virus. *Comptes rendus de l'Académie des Sciences* **228** (1949).

37. Burney, L. E. Influenza immunization: statement. *Public Health Report* **75,** 944 (1960).

38. Long, P. H. Recommendations for influenza immunization and control 1964-1965. *Medical Times* **92,** 1203–1205 (1964).

39. An, Q. *Models for statistical analyses of infectious disease data* PhD thesis (Emory University, 2014).

40. Grohskopf, L. A., Sokolow, L. Z., Broder, K. R., *et al.* Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices — United States, 2017–18 Influenza Season. *Morbidity and Mortality Weekly Report Recommendation Report* **66,** 1–20 (No. RR-2 2017).

41. Halloran, E. M., Haber, M., Longini, I. M. & Struchiner, C. J. Direct and indirect effects of vaccine efficacy and effectiveness. *American Journal of Epidemiology* **133,** 323–331 (1991).

42. Orenstein, W. A. *et al.* Field evaluation of vaccine efficacy. *Bulletin of the World Health Organization* **63,** 1055–1068 (1985).

43. Haber, M., Longini, I. M. & Halloran, M. E. Measures of the effects of vaccination in a randomly mixing population. *International Journal of Epidemiology* **20,** 300–310 (1991).

44. Smith, P. G., Rodrigues, L. C. & Fine, P. E. M. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *Int. J. Epidemiol.* **13,** 87–93 (1984).

45. 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. *Eurosurveillance* **18,** pii=20585 (2013).

46. Jackson, M. L. & Rothman, K. J. Effects of imperfect test sensitivity and specificity on observational studies of vaccine effectiveness. *Vaccine* **33,** 1313–1316 (2015).

47. *Evaluation of influenza vaccine effectiveness: a guide to the design and interpretation of observational studies* (World Health Organization, Geneva, 2017).

48. Ohmit, S. E. *et al.* Influenza vaccine effectiveness in the community and the household. *Clinical Infectious Diseases* **56,** 1363–1369 (2013).

49. Ohmit, S. E. *et al.* Influenza vaccine effectiveness in households with children during the 2012-2013 season: assessments of prior vaccination and serologic susceptibility. *Journal of Infectious Diseases* **211,** 1519–1528 (2015).

50. Cox, D. R. Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B* **43,** 187–220 (1972).

51. Petrie, J. G. *et al.* Influenza transmission in a cohort of households with children: 2010-2011. *PLoS One* **8** (2013).

52. Malosh, R. *et al.* Factors associated with influenza vaccine receipt in community dwelling adults and their children. *Vaccine* **32,** 1841–1847 (2014).

53. Monto, A. S. Studies of the community and family: acute respiratory illness and infection. *Epidemiologic Reviews* **16,** 351–373 (1994).

54. Smith, P. G. Assessment of the efficacy of BCG vaccination against tuberculosis using the case-control method. *Tubercle* **62,** 23–35 (1982).

55. Orenstein, E. W. *et al.* Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. *International Journal of Epidemiology* **36,** 623–631 (2007).

56. Skowronski, D. M. *et al.* Effectiveness of vaccine against medical consultation due to laboratory-confirmed influenza: results from a sentinel physician pilot project in British Columbia, 2004-2005. *Can Commun Dis Rep* **31,** 181–191 (2007).

57. 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,** 3104–3109 (2013).

58. Sullivan, S. G., Feng, S. & Cowling, B. J. Potential of the test-negative design for measuring influenza vaccine effectiveness: a systematic review. *Expert Rev. Vaccines* **13,** 1571–1591 (2014).

59. Broome, C. V., Facklam, R. R. & Fraser, D. W. Pneumococcal disease after pneumococcal vaccination: an alternative method to estimate the efficacy of pneumococcal vccine. *New England Journal of Medicine* **303,** 549–552 (1980).

60. Boom, J. A., Tate, J. E., Sahni, L. C., *et al.* Effectiveness of pentavalent rotavirus vaccine in a large urban population in the United States. *Pediatrics* **125,** e199–207 (2010).

61. Bar-Zeev, N., Kapanda, L., Tate, J. E., *et al.* Effectiveness of a monovalent rotavirus vaccine in infents in Malawi after programmatic roll-out: an observational and case-control study. *Lancet Infectious Diseases* **15,** 422–428 (2015).

62. Azman, A. S., Parker, L. A., Rumunu, J., *et al.* Effectiveness of one dose of oral cholera vaccine in response to an outbreak: a case-control study. *Lancet Global Health* **4,** e856–e863 (2008).

63. Franke, M. F., Jerome, J. G., Matias, W. R., *et al.* Comparison of two control groups for estimation of oral cholera vaccine effectiveness using a case-control study design. *Vaccine* **35,** 5812–5827 (2017).

64. Noronha, C. P., Struchiner, C. J. & Halloran, M. E. Assessment of the direct effectiveness of BC meningococcal vaccine in Rio De Janeiro, Brazil: a case-control study. *International Journal of Epidemiology* **24,** 1050–1057 (1995).

65. Skowronski, D. M. *et al.* 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,** 2842–2851 (2007).

66. Cowling, B. J. *et al.* Increased risk of non-influenza respiratory virus infections associated with receipt of inactivated influenza vaccine. *Clinical Infectious Diseases* **54,** 1778–1783 (2012).

67. Feng, S., Fowlkes, A. L., Steffens, A., Finelli, L. & Cowling, B. J. Assessment of virus interference in a test-negative study of influenza vaccine effectiveness. *Epidemiology* **28,** 514–524 (2017).

68. Bean, B. *et al.* Survival of influenza viruses on environmental surfaces. *Journal of Infectious Diseases* **146,** 47 (1982).

69. Morris, J. A. *et al.* Immunity to influenza to antibody levels. *New England Journal of Medicine* **274,** 527–535 (1966).

70. Gregg, M. B. The epidemiology of influenza in humans. *Annals of the New York Academy of Sciences* **353,** 45–53 (1980).

71. Fox, J. P. *et al.* Influenzavirus infections in Seattle families, 1975-1979. I. Study design, methods and the occurence of infections by time and age. *American Journal of Epidemiology* **116,** 212–227 (1982).

72. Johnson P. R., J. *et al.* Comparison of long-term systemic and secretory antibody respnses in children given live, attenuated, or inactivated influenza A vaccine. *Journal of Medical Virology* **17,** 325–335 (1985).

73. Fiore, A. E., Bridges, C. B. & Cox, N. J. in (eds Compans, R. W. & Orenstein, W. A.) chap. Seasonal Influenza Vaccines (Springer Berlin Heidelberg, 2009).

74. Carrat, F. *et al.* Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American Journal of Epidemiology* **167,** 775–785 (2008).

75. Johnson, P. R. *et al.* Immunity to influenza A virus infection in young children: a comparison of natural infection, live cold-adapted vaccine, and inactviated vaccine. *Journal of Infectious Diseases* **154,** 121–127 (1986).

76. Rocha, E. *et al.* Antigenic and genetic variation in influenza A (H1N1 virus isolates recovered from a persistently infected immunodeficient child. *Journal of Virology* **65,** 2340–2350 (1991).

77. Addy, C. L., Longini, I. M. & Haber, M. A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* **47,** 961–974 (1991).

78. Becker, N. G. The uses of epidemic models. *Biometrics* **35,** 295–305 (1979).

79. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London* **115,** 701–721 (1927).

80. Bailey, N. T. J. *The mathematical theory of epidemics* (Griffin, London, 1957).

81. Anderson, R. M. & May, R. M. *Infectious diseases of humans* (Oxford University Press, Oxford, 1957).

82. Keeling, M. J. & Eames, K. T. D. Network and epidemic models. *Journal of the Royal Society Interface* **2,** 295–307 (4 2005).

83. Grenfell, B. T. Chance and chaos in measles dynamics. *Journal of the Royal Statistical Society: Series B* **54,** 383–398 (1992).

84. Rohani, P., Earn, D. J. D. & Grenfell, B. T. The impact of immunisation on pertussis transmission in England and Wales. *Lancet* **355,** 285–286 (2000).

85. London, W. P. & Yorke, J. A. Recurrent outbreaks of measles, chickenpox, and mumps. *American Journal of Epidemiology* **98,** 453–468 (2008).

86. Longini, I. R., Ackerman, E. & Elveback, L. R. An optimization model for influenza A epidemics. *Mathematical Biosciences* **38,** 141–157 (1-2 1978).

87. Mills, C. E., Robins, J. M. & Lipsitch, M. Transmissibility of 1918 pandemic influenza. *Nature* **432,** 904–906 (2004).

88. Weaving, H. J., Rohani, P. & Keeling, M. J. Appropriate models for the management of infectious diseases. *PLoS Medicine* **2,** e174 (2005).

89. Brauer, F. in (eds Brauer, F., van den Driessche, P. & Wu, J.) chap. Modeling Influenza: Pandemics and Seasonal Epidemics (Springer, Berlin, Heidelberg, 2008).

90. Chowell, G., Miller, M. A. & Viboud, C. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiology and Infection* **136,** 852–864 (2008).

91. Longini, I. R. & Koopman, J. S. Household and community transmission parameters from final distributions of infections in households. *Biometrics* **38,** 115–126 (1982).

92. Halloran, E. M., Preziosi, M.-P. & Chu, H. Estimating vaccine efficacy from secondary attack rates. *JASA* **98,** 38–46 (2003).

93. Davis, X. M. & Haber, M. Estimating vaccine efficacy from household data observed over time. *Statistics in Medicine* **23,** 2961–2974 (2004).

94. Ohmit, S. E. *et al.* Influenza vaccine effectiveness in the 2011-2012 season: protection against each circulation virus and the effect of prior vaccination on estimates. *Clinical Infectious Diseases* **58** (2014).

95. McLean, H. Q. *et al.* Influenza vaccine effectiveness in the United States during 2012-2013: variable protection by age and virus type. *Journal of Infectious Diseases* **211,** 1529–1540 (2015).

96. Pierse, N. *et al.* Influenza vaccine effectiveness for hospital and community patients using control groups with and without non-influenza respiratory viruses detected, Auckland, New Zealand 2014. *Vaccine* **34,** 503–509 (2016).

97. Ohmit, S. E. *et al.* Substantial influenza vaccine effectiveness in households with children during the 2013-2014 influenza season, when 2009 pandemic influenza A(H1N1 virus predominated. *Journal of Infectious Diseases* **213,** 1229–1236 (2016).

98. Haber, M. *et al.* A probability model for evaluating the bias and precision of influenza vaccine effectiveness estimates from case-control studies. *Epidemiology and Infection* **143,** 1417–1426 (2015).

99. Shi, M., An, Q., Ainslie, K. E. C., Haber, M. & Orenstein, W. A. A comparison of the test-negative and the traditional case-control study designs for estimation of influenza vaccine effectiveness under nonrandom vaccination. *BMC Infectious Diseases* **17,** 757–777 (2017).

100. Fukushima, W. & Hirota, Y. Basic principles of test-negative design in evaluating influenza vaccine effectiveness. *Vaccine* **35,** 4796–4800 (2017).

101. Kissling, E. *et al.* "I-MOVE" towards monitoring seasonal and pandemic influenza vaccine effectiveness: lessons learnt from a pilot multi-centric case-control study in Europe, 2008-9. *Eurosurveillance* **14** (2009).

102. Treanor, J. J. *et al.* Effectiveness of seasonal influenza vaccines in the United States during a season with circulation of all three vaccine strains. *Clinical Infectious Diseases* **55,** 951–959 (2012).

103. Skowronski, D. M. *et al.* Interim estimates of influenza vaccine effectiveness in 2012/13 from Canada's sentinel surveillance network, January 2013. *Eurosurveillance* **18** (2013).

104. Turner, N. *et al.* Interim estimates of the effectiveness of seasonal trivalent inactivated influenza vaccine in preventing influenza hospitalisations and primary care visits in Auckland, New Zealand, in 2014. *Eurosurveillance* **19** (2014).

105. Carville, K. S. *et al.* Understanding influenza vaccine protection in the community: an assessment of the 2013 influenza season in Victoria Australia. *Vaccine* **33,** 341–345 (2015).

106. Jackson, M. L. & Nelson, J. C. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine* **17,** 2165–2168 (2013).

107. Centers for Disease Control and Prevention. Prevention and control of influenza with vaccines: recommendations from the Advisory Committee on Immunization Practices (ACIP) - United States, 2013-2014. *Morbidity and Mortality Weekly Report* **62** (2013).

108. Haber, M., Longini, I. M. & Cotsonis, G. Models for the statistical analysis of infectious disease data. *Biometrics* **44,** 163–173 (1988).

109. Davis, X. M. & Haber, M. *Estimation of vaccine efficacy from household data* CD-ROM. 2001.

110. Halloran, M. E. in (eds Thomas, J. C. & Weber, D. J.) chap. Concepts of Transmission and Dynamics (Oxford University Press, New York, NY, 2001).

111. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing* **16,** 1190–1208 (1995).

112. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2016). https://www.R-project.org/.

113. U.S. Census Bureau. *Cesus 2010, Summary File 1, Households and Families: 2010 and QT-P11* http://factfinder.census.gov (2016).

114. Hayward, A. C. *et al.* Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. *Lancet Respir Med* **2,** 445–454 (2014).

115. Leung, N. H. L., Xu, C., Ip, D. K. M. & Cowling, B. J. The fraction of influenza virus infections that are asymptomatic: a systematic review and meta-analysis. *Epidemiology* **26,** 862–872 (2015).

116. Foy, H. M., Cooney, M. K., Allan, I. D. & Albrecht, J. K. Influenza B in households: virus shedding without symptoms or antibody response. *American Journal of Epidemiology* **126,** 506–515 (3 1987).

117. Wallinga, J., Teunis, P. & Kretzschmar, M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* **164,** 936–944 (2006).

118. Mossong, J. *et al.* Social contacts and mixing patters relevant to the spread of infectious diseases. *PLoS Medicine* **5,** 381–391 (2008).

119. Castilla, J. *et al.* Influenza vaccine effectiveness in preventing outpatient, inpatient, and severe cases of laboratory-confirmed influenza. *Clinical Infectious Diseases* **57,** 167–175 (2013).

120. Deiss, R. *et al.* Vaccine-associated reduction in symptom severity among patients with influenza A/H3N2 disease. *Vaccine* **33,** 7160–7167 (2015).

121. 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 Infectious Diseases* **14,** 231 (2014).

122. Sullivan, S. G., Tchetgen, T. & Cowling, B. J. Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *American Journal of Epidemiology.* **184,** 345–353 (2016).

123. Lipsitch, M., Jha, A. & Simonsen, L. Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. *International Journal of Epidemiology* **45,** 2060–2074 (2016).

124. Ainslie, K. E. C., Shi, M., Haber, M. & Orenstein, W. A. On the bias of estimates of influenza vaccine effectiveness from the test-negatvie studies. *Vaccine* **35,** 7297–7301 (2017).

125. Dawood, F. S. *et al.* Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet Infectious Diseases* **12,** 687–695 (2012).

126. Ramsay, L. C. *et al.* The impact of repeated vaccination on influenza vaccine effectiveness: a systematic review and meta-analysis. *BMC Medicine* **15,** 159–176 (2017).

127. Hoskins, T., Davies, J., Smith, A., Miller, C. & Allchin, A. Assessment of inactivated influenza A vaccine after three outbreaks of influenza A at Christ's Hospital. *Lancet* **313,** 33–35 (1979).

128. Keitel, W. A., Cate, T. R., Huggins, R. B.C.L. L. & Hess, K. R. Efficacy of repeated annual immunization with inactivated influenza virus vaccines over a five year period. *Vaccine* **15,** 1114–1122 (1997).

129. Beyer, W. E., de Bruijin, I. A., Palache, A. M., Westendorp, R. G. & Osterhaus, A. D. Protection against influenza after annually repeated vaccination: a meta-analysis of serologic and field studies. *Archives of Internal Medicine* **15,** 182–188 (1999).

130. McLean, H. Q. *et al.* Impact of repeated vaccination on vaccine effectiveness against influenza A (H3N2) and B during 8 seasons. *Clinical Infectious Diseases* **59,** 1375–1385 (2014).

131. Skowronski, D. M. *et al.* A perfect storm: impact of genomic variation and serial vaccination on low influenza vaccine effectiveness during 2014-2015 season. *Clinical Infectious Diseases* **63,** 21–32 (2016).

132. Hancock, K., Veguilla, V., Lu, X., *et al.* Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *New England Journal of Medicine* **361,** 1945–1952 (2009).

133. Miller, E., Hoschler, K., Hardelid, P., *et al.* Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet* **375,** 1100–1108 (2010).

134. Castilla, J. *et al.* Decline in influenza vaccine effectiveness with time after vaccination, Navarre, Spain, season 2011/12. *Eurosurveillance* **18,** pii=20388 (2013).

135. Jiménez-Jorge, S., de Mateo, S., Delgado-Sanz, C., *et al.* Effectiveness of influenza vaccine against laboratory-confirmed influenza, in the late 2011-2012 season in Spain, among population targeted for vaccination. *BMC Infectious Diseases* **13,** 441 (2013).

136. Kissling, E., Valenciano, M., Larrauri, A., *et al.* Low and decreasing vaccine effectiveness against influenza A(H3) in 2011/12 among vaccination target groups in Europe: results from the I-MOVE multicentre case-control study. *Eurosurveillance* **18,** pii=20390 (2013).

137. Pebody, R., Andrews, N., McMenamin, J., *et al.* Vaccine effectiveness of 2011/12 trivalent seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: evidence of waning intra-seasonal protection. *Eurosurveillance* **18,** pii=20389 (2013).

138. Sullivan, S. G. *et al.* Influenza vaccine effectiveness during the 2012 influenza season in Victoria, Australia: influences of waning immunity and vaccine match. *Journal of Medical Virology* **86,** 1017–1025 (2014).

139. Andrews, N., McMenamin, J., Durnall, H., *et al.* Effectiveness of trivalent seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: 2012/13 end of season results. *Eurosurveillance* **19,** 5–13 (2014).

140. Kissling, E., Nunes, B., Robertson, C., others & Team, I.-M.C. S. I-MOVE multicenter case-control study 2010/11 to 2014/15: Is there within- season waning of influenza type/subtype vaccine effectiveness with increasing time since vaccination? *Eurosurveillance* **21,** pii=30201 (2016).

141. Ferdinands, J. M. *et al.* Intraseason Waning of Influenza Vaccine Protection: Evidence From the US Influenza Vaccine Effectiveness Network, 2011–2012 Through 2014–2015. *Clinical Infectious Diseases* **64,** 544–550 (2017).

142. Reed, C. *et al.* Estimates of the prevalence of (H1N1) 2009, United States, April-July 2009. *Emerging Infectious Diseases* **15,** 2004–2007 (2009).