Using Double Negative Controls to Adjust for Healthy User Bias in a Recombinant Zoster Vaccine Safety Study

Iris Emerman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2021

Reading Committee:

Jennifer Nelson, Chair

Xu Shi

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Using Double Negative Controls to Adjust for Healthy User Bias in a Recombinant Zoster Vaccine Safety Study

Iris Emerman

Chair of the Supervisory Committee:
Affiliate Professor Jennifer Nelson
Department of Biostatistics

Bias due to unmeasured confounding haunts the field of pharmacosurveillance. We applied a novel causal inference method utilizing negative control variables to correct for unmeasured confounding bias in a safety study of Recombinant Zoster Vaccine. This ‘double negative control’ method uses a confounding bridge to estimate the magnitude and scale of bias, relying on weaker assumptions than common causal inference methods. A principled approach was developed to determine negative control outcome and exposure pairs capable of correcting for healthy user bias. Relative risk estimates and confidence intervals calculated using double negative control methodology were compared to those calculated using a flexible propensity score regression method. The double negative control methodology was feasible to implement in an electronic health record setting, adjusted estimates in the expected direction (e.g., attenuated protective effects that were potentially due to healthy user bias), and more accurately quantified uncertainty due to unmeasured confounding.

TABLE OF CONTENTS

List of Figures iii

List of Tables iv

Glossary v

Chapter 1. Introduction 7

Chapter 2. Methods 12

2.1 Study design 12

2.1 Exposure and comparator groups 13

2.2 Safety outcomes of interest 14

2.3 Covariates 15

2.4 Overview of statistical analysis 16

2.4.1 Flexible propensity score-adjusted regression analysis 16

2.4.2 Selecting negative control outcome and exposure pairs 17

2.4.3 Double negative control assumptions and details 20

2.5 Sensitivity analysis 21

2.5.1 Analysis using well-visit + flu comparator 21

2.5.2 Comparing across double negative controls 22

Chapter 3. Results 22

3.1 Selecting double negative controls 22

3.2 Baseline characteristics 23

3.3 Comparing relative risk estimates for safety outcomes across methods 25

3.1 Sensitivity analysis 26

3.1.1 Analysis using well-visit + flu comparator 26

3.1.2 Comparing estimates across double negative controls 27

Chapter 4. Discussion 28

References 35

Appendix A: Supplementary tables 42

Appendix B: Review of select methods to account for unmeasured confounding 45

Appendix C: Summary of key assumptions for nonparametric identification in a double negative control design 47

LIST OF FIGURES

Figure 1: Directed acyclic graph (DAG) of an example double negative control. 9

Figure 2: Flow chart of principled approach to determine primary double negative controls. 18

Figure 3: Relative risk of acute myocardial infarction between RZV vaccinees and well-visit comparators estimated using different double negative controls. 28

LIST OF TABLES

Table 1: Operational definitions of safety outcomes. 15
Table 2: Operational definitions of negative control variables used in primary analysis. 23
Table 3: Characteristics of the study population..... 24
Table 4: Relative risk of target safety outcomes between RZV vaccinees and well-visit comparators. 26
Table 5: Relative risk of target safety outcomes between RZV vaccinees and well-visit + flu comparators. 27
Table 6: Operational definitions of covariates used in analysis. 42
Table 7: Operational definitions of all negative control variables considered for potential use in analysis. 42
Table 8: Relative risk of target safety outcomes RZV vaccinees and well-visit comparators estimated using different double negative controls..... 44

GLOSSARY

CI: Confidence interval

DAG: Directed acyclic graph

ED: Emergency department

GERD: Gastroesophageal reflux disease

ICD-10-CM: International Classification of Diseases, Tenth Edition - Clinical Modification

IP: Inpatient

IPTW: Inverse probability of treatment weighting

IV: Instrumental variable

MCO: Managed care organizations

MI: Myocardial infarct

NCE: Negative control exposure

NCO: Negative control outcome

OP: Outpatient

OPH: Ophthalmology

OPT: Optometry

RR: Relative risk

RZV: Recombinant Zoster Vaccine

UC: Urgent Care

VSD: Vaccine Safety Datalink

ZVL: Zoster Vaccine Live

ACKNOWLEDGEMENTS

I would like to thank Dr. Jennifer Nelson and Dr. Xu Shi for their guidance, encouragement, and mentorship throughout the process of writing this dissertation. Xu, putting your methods into practice was a fantastic learning experience that I will never forget. I would like to thank Ernesto Ulloa Pérez for your advice and for allowing me the use of your code – I will look at it whenever I need inspiration. I would like to thank Gitana Garofalo for your enthusiasm and sparkling conversation. Finally, I would like to thank my parents and my brother, for supporting me in everything I do.

Chapter 1. INTRODUCTION

Confounding bias in observational studies often emerges when groups being compared differ in ways that affect the outcome of interest but are not accounted for in the statistical analysis. One potential source of confounding bias is the healthy user effect. The healthy user effect is the propensity for patients receiving one preventative service to engage in other health-promoting behaviors as well.[1] Take, for example, an observational study on the effect of estrogen replacement therapy on risk of coronary heart disease. If relevant health-promoting behaviors such as healthy diet and exercise were not controlled for, then the protective effect of the treatment on risk of coronary heart disease may be overestimated.[1,2]

There is no simple way to control for healthy user bias in observational studies. Many existing methods used to adjust for healthy user bias assume that health-seeking behaviors can be accurately measured and that, once measured factors are adjusted for, there is no unmeasured confounding.[3] A common approach involves measuring proxies for health-seeking behaviors and then applying traditional epidemiological methods to analytically adjust for those measured proxies. These methods include outcome regression, inverse probability of treatment weighting (IPTW), matching, stratification, and restriction of the comparator group.[3-13] In cases where proxy variables are unmeasured or inadequate, studies may vastly overestimate the protective effect of preventative treatments.[14] For example, influenza typically accounts for fewer than 10% of all deaths among the elderly each winter, but observational studies have reported that influenza vaccination reduces all-cause winter mortality among the elderly by up to 50%, clearly demonstrating that substantial bias is at play.[15,16]

When confounding factors cannot all be accurately measured and incorporated into analyses, methods that account for unmeasured confounding are needed. One such approach that is commonly used is the use of instrumental variables (IV).[17-22] An IV is a pre-treatment variable that satisfies the following three assumptions: 1) the IV is associated with the treatment of interest; 2) the IV does not directly affect the outcome except as mediated by the treatment of interest; and 3) the IV is independent of the unmeasured confounders.[23] For example, physician-specific prescribing preference can be used as an IV in the context of pharmacosurveillance.[17] Causal effects can be estimated with the addition of a fourth assumption, such as the monotonicity assumption (i.e., no patients for whom a physician who prefers treatment A would have prescribed treatment B and vice versa) or the assumption of no current treatment interaction (i.e., physician-specific prescribing preference does not modify the effect of treatment on outcome).[19,24,25] As detailed in Appendix C, a valid IV is also a valid negative control exposure (NCE). A selection of additional alternative methods designed to account for unmeasured confounding is detailed in Appendix B.

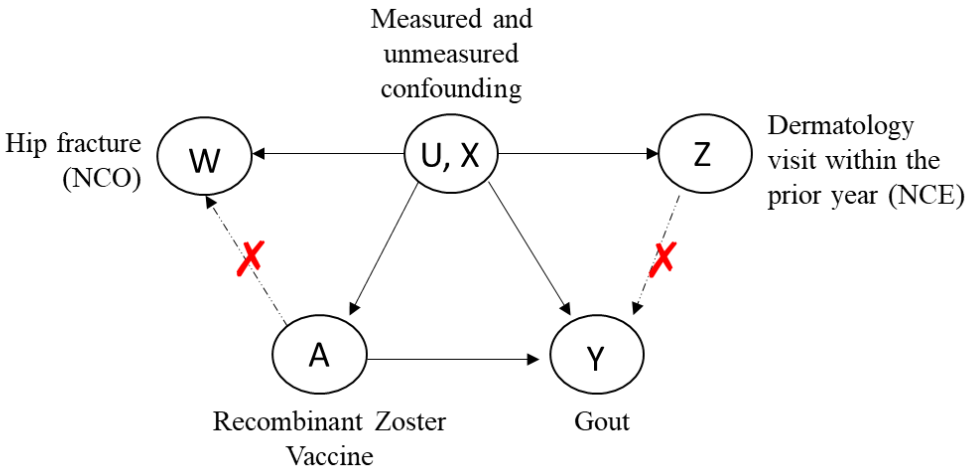

Figure 1: Directed acyclic graph (DAG) of an example double negative control. This DAG illustrates the hypothetical causal effect of Recombinant Zoster Vaccine (*A*) on gout (*Y*) subject to confounding by measured confounders (*X*) and unmeasured confounders (*U*). In this example, the negative control exposure (NCE; *Z*) is a dermatology visit within the prior year and the negative control outcome (NCO; *W*) is hip fracture. Both *Z* and *W* are proxies for *U*, healthy-seeking behavior.

Another popular approach that is designed to address unmeasured confounding is the use of negative control variables. Negative control methods can be used to detect, reduce, or correct for unmeasured confounding through an alternative utilization of the measured proxies of unmeasured confounders. Negative control outcomes (NCOs) are proxies of the unmeasured confounders that are not causally affected by the treatment of interest. NCEs are proxies of the unmeasured confounders that do not causally affect the outcome of interest. Figure 1 illustrates these associations in a causal diagram. In current practice, studies using NCOs and NCEs often focus on bias detection.[26] Bias detection methods typically test for an association between target outcome or exposure and a single negative control variable.[14,27-31] If there were no bias present, then there would be no effect of an NCE (e.g., dental care participation) on the outcome (e.g., breast cancer mortality).[27] For example, in an influenza vaccine study, researchers used the NCO of all-cause mortality during the time period prior to influenza season (i.e., a time when influenza

was not yet circulating so no effect of vaccine existed) to quantify the size of the healthy user bias potentially impacting estimates of the effect of influenza vaccination on the primary outcome of interest: mortality *during* influenza season.[14]

Recently, there has been increasing interest in the use of negative control variables to reduce or remove unmeasured confounding bias, particularly in the context of pharmacosurveillence.[32-43] Empirical distribution calibration is an example of a negative control method designed to reduce bias. Empirical distribution calibration involves calculating effect estimates for associations between a primary treatment and NCOs or between a primary outcome and NCEs. The null hypothesis of no association between primary and negative control variables is true by definition, so a probability distribution can be fitted to those effect estimates to create an empirical null distribution.[32,33] P-values and confidence intervals can be calibrated using this null distribution, reducing bias caused by unmeasured confounding by taking into account residual systematic error. An example of a method that goes a step further, eliminating unmeasured confounding bias through negative controls, is a recently developed method that adjusts for unobserved confounding in the Cox proportional hazards regression model by subtracting the estimated log-hazards ratio for the association between treatment and NCO from that of the treatment and target outcome.[37] This method relies on the monotonicity assumption and the assumption that primary and negative outcomes are time-to-event outcomes following a certain Cox proportional hazards model. Another example is the difference-in-difference (DID) approach, which can be formally justified as a method of using an NCO to correct for bias.[40] DID is typically used to estimate the effect of a treatment of interest by comparing changes in outcomes over time between treatment groups. The NCO, a baseline measure of the outcome of interest prior to the treatment, can be used to correct for bias because we assume ‘bias

equivalence.'[40] When bias equivalence is not a plausible assumption (i.e., the outcome of interest and the NCO are not measured on comparable scales), a less restrictive option should be used. A scale-invariant generalization of the DID was recently proposed which relaxes the assumption of bias equivalence.[40]

Double negative controls, defined as NCE and NCO pairs, have been proposed as a new way to nonparametrically identify causal effects by estimating the scale and magnitude of bias.[42-44] By using both an NCE and an NCO, double negative control methods are able to relax some of the assumptions, such as the monotonicity assumption, that were necessary in previous negative control methods designed to remove unmeasured confounding bias. Key to the identification of causal effects in double negative control methodology is the 'confounding bridge' function, which links the potential outcome mean and on the NCO distribution.[44] The confounding bridge function characterizes the relationship between the confounding effect of the unmeasured confounder on the target outcome and the NCO. Identifiable using an NCE under certain completeness conditions, the confounding bridge function can be used for the estimation of causal effect in the presence of unmeasured confounding. Double negative control methods have been generalized to a framework called 'proximal causal inference,' which acknowledges that covariate measurements can be imperfect proxies and partitions the covariates into three types to improve confounding adjustment, expanding the types of variables that can be used as independent proxies in the identification of causal effects.[45,46]

This thesis demonstrates the novel application of double negative control methodology to correct for healthy user bias due to unmeasured confounding in the estimation of the effect of Recombinant Zoster Vaccine (RZV) on selected safety outcomes of interest. In addition, this work begins to develop a standardized approach for implementation of this methodology in

practice and specifically within an electronic health data study setting. RZV is a 2-dose vaccine intended to prevent herpes zoster, also known as shingles, and its complications (Shingrix®, GlaxoSmithKline). It is recommended by the U.S. Advisory Committee on Immunization Practices (ACIP) for immunocompetent adults aged 50 and above, including those who have received Zoster Vaccine Live (ZVL), a vaccine previously recommended by the ACIP for herpes zoster prevention.[47,48] Prior research has shown that estimates of the effects of vaccines in adults, especially older adults, are highly susceptible to the healthy user effect.[14] This is because patients who receive voluntary vaccinations such as influenza or pneumococcal vaccinations tend to also engage in other health-promoting behaviors.[49] Patients who receive RZV, like those who receive other voluntary vaccines, are expected to demonstrate healthy user effects, making adjustment for healthy user bias a vital step in the assessment of RZV safety outcomes.

Chapter 2. METHODS

2.1 STUDY DESIGN

This research is built off a prospective post-licensure RZV safety study conducted on the same data source.[50] Data acquisition was made possible by the Vaccine Safety Datalink (VSD), a collaboration between nine integrated healthcare systems and the Centers for Disease Control and Prevention (CDC). Electronic health record data on immunizations, medical utilizations, and demographics were captured on men and women aged 50 and older who either received RZV or had a qualifying annual well visit at one of seven VSD-participating managed care organizations (MCO): Kaiser Permanente of Northern California, Oakland, California; Kaiser Permanente of Southern California, Pasadena, California; Marshfield Clinic Research Foundation; Marshfield, Wisconsin; and Kaiser Permanente Northwest, Portland, Oregon. The study period began in

January 2018, when RZV usage began at some VSD sites, and continued until December 2019. Each MCO's institutional review board approved this study; informed consent was not required.

2.1 EXPOSURE AND COMPARATOR GROUPS

The primary exposure group for this analysis was composed of adults aged 50 and older who received dose 1 and/or dose 2 of RZV (representing >99% of all doses RZV vaccine doses received during study period), as identified by Administered Vaccine (CVX) code 187, at the participating MCOs during the surveillance period. Primary analysis compared the RZV group with adults aged 50 and older who, during the study period, sought care at the participating MCOs for an annual well-visit and did not receive RZV on or before the well-visit date. Annual well-visits were identified by ICD-10-CM diagnosis codes Z00.00 and Z00.01. For simplicity, we refer to this group as 'well-visit comparators.' If an individual had more than one well-visit during the surveillance period, one representative visit was selected per study year (2018 and 2019). Individuals were excluded from the well-visit comparator group if they had unknown gender information or had more than three well-visits during the surveillance period (as this unusually high frequency could either be an error or reflective of unusual health-seeking behavior). In sensitivity analyses, we used a subset of the well-visit group composed of all participants in this group who also had received an influenza vaccine in the year prior to their well-visit. This comparator group, which we hereafter refer to as the 'well-visit + flu' subset for simplicity, was the main comparator group used in the original RZV safety study.[50]

We chose well-visit comparators instead of the well-visit + flu subset as our primary comparator group because the differences in healthy user behaviors between RZV recipients and well-visit comparators were greater than the differences between RZV recipients and the well-visit + flu subset in the original RZV safety study.[50] Specifically, well-visit comparators

exhibited lower frequencies of health-seeking behaviors (e.g., having a optometry or ophthalmology visit in the prior year) than RZV recipients. Given this, we hypothesized that healthy user bias would be larger when making comparisons with the well-visit group, thus enhancing our ability to demonstrate potential correction for such bias using the double negative control methodology. In other words, we hypothesized that the incremental benefit of our double negative control analysis on reducing bias in relative risk (RR) estimates when compared to estimates obtained via the flexible propensity score regression analysis used in the original RZV safety surveillance study would be larger when using well-visit comparators (our primary analyses) versus the well-visit + flu subset (our sensitivity analysis).

2.2 SAFETY OUTCOMES OF INTEREST

Safety outcomes were chosen according to clinical significance, potential for unmeasured confounding due to healthy user bias, and feasibility. Relevance of an outcome was determined based on safety profiles of prior vaccines as well as data from pre-licensure studies of RZV.[51] The unmeasured confounder that was most concerning was the healthy user effect, so we prioritized safety outcomes that were determined, through a combination of expert opinion and prior research, to have greater susceptibility to healthy user bias. With respect to feasibility, the causal inference method used in this analysis has not been fully examined in settings with extremely rare outcomes, so the rarest safety outcomes were avoided for the purposes of this analysis.[42] Safety outcomes examined in this analysis were stroke, acute myocardial infarction (MI), pneumonia, and gout. Outcomes were defined using ICD-10-CM diagnoses assigned by health care providers to outpatient (OP), emergency department (ED), urgent care (UC), or inpatient (IP) hospital encounters. Each outcome was assessed on Day 1 (i.e., the day after the index RZV vaccination or well-visit date) through Day 42. Outcomes were excluded if any of the

ICD-10-CM codes specifying that outcome were recorded in the 365 days prior to the index date. This was done to avoid inclusion of follow-up appointments for events that occurred prior to the index. Table 1 details the complete list and explicit definitions of all safety outcomes.

Table 1: Operational definitions of safety outcomes.

Safety Outcomes	ICD-10 Code(s)[a]	Medical Setting	Primary double negative control
Acute myocardial infarction	I21	IP, ED/UC, OP	NCE: Prior history of ZVL NCO: Hip fracture
Stroke	I63, I61.9	IP, ED/UC, OP	NCE: Dermatology visit NCO: Epistaxis
Gout	M10.	IP, ED/UC, OP	NCE: Dermatology visit NCO: Hip fracture
Pneumonia	J18, B01.2, J11.0, J12, J13, J14, J15, J16	IP, ED/UC, OP	NCE: Prior history of ZVL NCO: Hip fracture

[a]3-digit codes (e.g., 345) included those that started with those 3 digits and contained any additional 4th or 5th digits (e.g., 345.11); 4-digit codes included those that started with those 4 digits and had any 5th digit.

2.3 COVARIATES

Baseline characteristics were captured in the 365 days prior to the index RZV vaccination or well-visit date. Selected covariates included age, MCO site, gender, indicators of comorbid conditions, and indicators of health care utilization that may reflect health-seeking behavior. Comorbid conditions included diabetes, hypertension, hyperlipidemia, ischemic conditions (ischemic heart disease, transient ischemic attack, or prior stroke), gastroesophageal reflux disease (GERD), osteoarthritis, atrial fibrillation, herpes zoster, dementia, congestive heart failure (CHF), and chronic obstructive pulmonary disease (COPD). Indicators of health care utilization included dermatology visit, optometry/ophthalmology visit, receipt of influenza vaccine in the prior year, and any prior history of ZVL vaccination. Table 6 in Appendix A

details all covariates included in the analysis that were defined using ICD-10-CM, CVX, or department codes.

2.4 Overview of statistical analysis

To assess the added utility of the double negative control methodology, we compared it to the same general statistical analysis approach that was used in the original RZV safety surveillance study using this data source. The propensity score method used in the original RZV safety study attempts to controls for healthy user bias by adjusting for measured proxies of health-seeking behavior. Unlike double negative control methods, it assumes that the proxies are perfect measures of the unmeasured confounders and there is no residual unmeasured confounding. [50,52] Analysis using double negative control methodology incorporated the same health-seeking behavior proxy variables to control for unmeasured confounding in a different way, acknowledging that the proxies were imperfect measures of the unmeasured confounders.

2.4.1 *Flexible propensity score-adjusted regression analysis*

As in the original RZV safety surveillance study using this data source, we used a flexible propensity score regression method to estimate marginal RR and 95% confidence intervals for each target outcome among RZV recipients compared with well-visit (or well-visit + flu) comparators.[50] The marginal RR of each target outcome was estimated using the following steps: 1) We estimated the probability of receiving RZV (i.e., the propensity score) given potential measured confounders. For non-cardiovascular outcomes (i.e., pneumonia, gout), the covariates used were age group, gender, and select healthy-user indicators (receipt of influenza vaccine in the prior year, ophthalmology/ophthalmology visit in the prior year, dermatology visit in the prior year, and any prior history of ZVL). For cardiovascular related outcomes (i.e., stroke, acute

MI), we added the covariates hypertension, hyperlipidemia, diabetes, and ischemic heart condition (presence of ischemic heart disease, transient ischemic attack, or prior stroke); 2) We used logistic regression to model the association between each outcome and receipt of RZV vaccine. We used splines to flexibly adjust for the propensity score; 3) We used standardization to obtain average, population-level risks for RZV and well-visit groups. Standardization is a causal inference technique which predicts, for each participant, what their outcome risk would be if they were in the RZV group and what their risk would be if they were in the well-visit group.[53] Empirical averages were then computed for each exposure group (e.g., RZV and well-visit); 4) We computed the marginal RR as the ratio of these two averages and computed corresponding 95% confidence intervals under the theory of M-estimation.[52,54]

2.4.2 *Selecting negative control outcome and exposure pairs*

For simplicity, let *A*, *Y*, and *X* denote primary exposure (RZV or well-visit), outcome (e.g., stroke), and a vector of observed covariates (e.g., age and gender), respectively. Let *Z*, *W*, and *U* denote negative control exposure (e.g., dermatology visit), negative control outcome (e.g., hip fracture), and unmeasured confounder (i.e., healthy user effect), respectively. Throughout this thesis, *Z* will be used interchangeably with NCE and *W* will be used interchangeably with NCO.

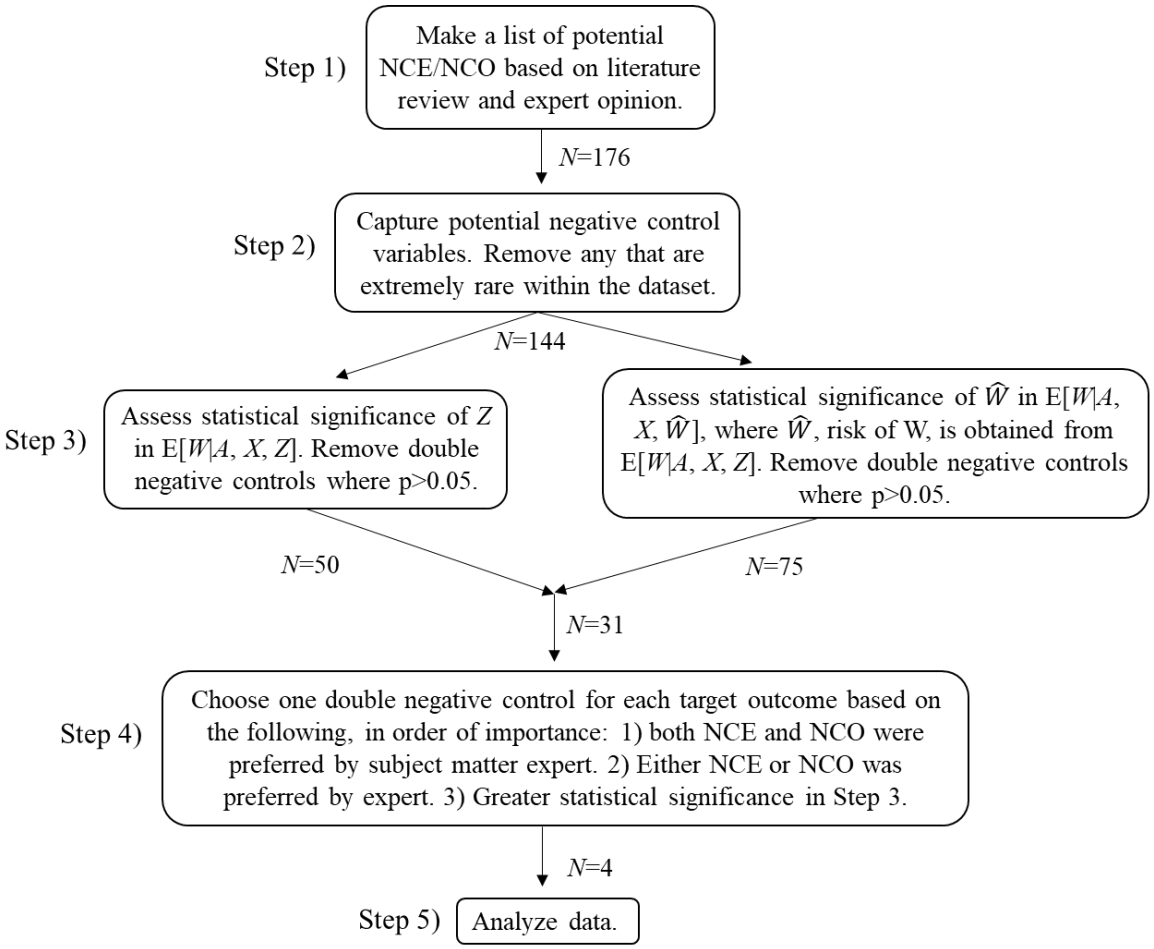

Figure 2: Flow chart of principled approach to determine primary double negative controls. Notation: *W*, negative control outcome; *Z*, negative control exposure, *A*; exposure of interest; *Y*, outcome of interest; *X*, measured confounders; *N*, remaining double negative controls under consideration.

The process we used to determine double negative controls for each of the four target outcomes of interest is outlined in Figure 2. Mathematical assumptions underlying this process are described in Appendix C.[41,42,52] A wide range of potential negative controls variables were initially chosen based on a combination of literature review and subject matter expert knowledge. Potential NCOs were associated with health-seeking behavior given the measured confounders but were not plausibly caused by either the treatment of interest (RZV) or by the NCE. Potential

NCEs were associated with health-seeking behavior given the measured confounders and the treatment of interest but were not causally related to any of the four target outcomes.

NCEs collected for potential use in this analysis were: receipt of influenza vaccine,[49] prior history of ZVL,[50] dermatology visit,[50] optometry/ophthalmology visit,[50] colon cancer screening,[49,55] cataract,[56,57] lipomas,[56-58] ingrown nail,[56-58] eyelid disorders,[56,57] and impacted cerumen.[58] All potential NCEs, other than receipt of ZVL, were collected in the year prior to the index RZV vaccination or well-visit date. Prior history of ZVL was defined as any prior record of ZVL vaccination on or before the index RZV vaccination or well-visit date. NCOs collected for potential use in this analysis were: hip fracture,[14,56,57,59] wrist fracture,[56-58] epistaxis,[56,57] and leg cellulitis.[58] All potential NCOs were collected in the 42 days after the index RZV vaccination or well-visit date

Once potential negative controls were identified, they were operationally defined using electronic health record data, assessed for usability within the dataset, and then reduced to a usable set. First, the frequency of each NCO and NCE was calculated in the study cohort. Potential NCEs that were exceedingly rare relative to RZV (which was present in 25.8% of participants) were eliminated as potential NCEs. Potential NCOs that were exceedingly rare relative to the four outcomes of interest (whose frequencies ranged from 0.044% to 0.26%) were eliminated as potential NCOs. Next, each potential double negative control was evaluated in a series of steps to ensure that it provided sufficient information about the unmeasured confounding in the dataset. First, the statistical significance of Z in $\mathrm{E}[W \mid A, X, Z]$ was calculated to check whether Z and W were associated due to the latent Z-U-W pathway. Then, the statistical significance of $\widehat{W}$ in $\mathrm{E}[Y \mid A, X, \widehat{W}]$, where $\widehat{W}$ was the risk of negative control outcome predicted from $\mathrm{E}[W \mid A, X, Z]$, was calculated to check in the confounding bridge was well-estimated.

Finally, if the statistical significance of either was greater than 0.05, the double negative control was eliminated as a potential option for use in our analysis.

One primary double negative control was chosen for each outcome of interest in our analysis. The primary double negative control was chosen from the remaining candidates based on the following, in order of importance: 1) both NCE and NCO were preferred by the subject matter expert; 2) either NCE or NCO were preferred by the subject matter expert; and 3) greater marginal significance of Z in E[$W \mid A, X, Z$] and of $\widehat{W}$ in E[$Y \mid A, X, \widehat{W}$]. Negative control variables preferred by the subject matter expert were those variables that had previously been used by the authors as healthy user indicators or as negative controls. These variables were influenza vaccine, dermatology visit, optometry/ophthalmology visit, prior history of ZVL, and hip fracture.

2.4.3 *Double negative control assumptions and details*

In the presence of an unmeasured confounder U, an observed difference in the outcome between the treatment and control groups is a combination of confounding bias and underlying causal effect. Since U is not measured, it is not possible to directly differentiate the amount of variation in the target outcome due to the treatment of interest from the amount of variation due to U. The goal of using double negative controls is to leverage available proxies of U in order to indirectly remove the variation due to U.[42-44] By definition, any difference in W between the treatment group (A=1) and comparator group (A=0) can only be attributed to U. This is referred to as the "confounding effect on the negative control outcome."[44] Similarly, any difference in Y (or W) between the Z=1 group and Z=0 group can only be attributed to U after controlling for A.

Logistic regression was used to model both the target outcome and the negative control outcome. For non-cardiovascular outcomes (i.e., pneumonia, gout), the covariates used were age

group, gender, and healthy-user indicators (receipt of influenza vaccine in the prior year, ophthalmology/ophthalmology visit in the prior year, dermatology visit in the prior year, and prior history of ZVL) that were not being used as the NCE for that target outcome. For cardiovascular related outcomes (i.e., stroke, acute MI), we added the covariates hypertension, hyperlipidemia, diabetes, and ischemic heart condition (presence of ischemic heart disease, transient ischemic attack, or prior stroke).

$$\begin{bmatrix} E(Y|A=1,Z=1,\boldsymbol{X}) & E(Y|A=1,Z=0,\boldsymbol{X}) \end{bmatrix} \begin{bmatrix} \Pr(W=1|A=1,Z=1,\boldsymbol{X}) & \Pr(W=1|A=1,Z=0,\boldsymbol{X}) \\ \Pr(W=0|A=1,Z=1,\boldsymbol{X}) & \Pr(W=0|A=1,Z=0,\boldsymbol{X}) \end{bmatrix}^{-1} \begin{bmatrix} \Pr(W=1|\boldsymbol{X}) \\ \Pr(W=0|\boldsymbol{X}) \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} E(Y|A=0,Z=1,\boldsymbol{X}) & E(Y|A=0,Z=0,\boldsymbol{X}) \end{bmatrix} \begin{bmatrix} \Pr(W=1|A=0,Z=1,\boldsymbol{X}) & \Pr(W=1|A=0,Z=0,\boldsymbol{X}) \\ \Pr(W=0|A=0,Z=1,\boldsymbol{X}) & \Pr(W=0|A=0,Z=0,\boldsymbol{X}) \end{bmatrix}^{-1} \begin{bmatrix} \Pr(W=1|\boldsymbol{X}) \\ \Pr(W=0|\boldsymbol{X}) \end{bmatrix} \quad (2)$$

Y(a) denotes the potential outcome had the individual received treatment (a=1) or not (a=0). $E[Y(1)]$, the average risk of potential outcome $Y(1)$, was estimated using Equation (1) and $E[Y(0)]$, the average risk of potential outcome $Y(0)$, was estimated using Equation (2). Although our covariates were discrete, these equations also apply if one or more covariates are continuous. We computed the marginal RR as the ratio of mean potential outcome estimates. To calculate standard error, we computed the sandwich variance using M-estimation methodology, incorporating the uncertainty from all intermediate modeling steps in the calculation of RR.[54]

2.5 SENSITIVITY ANALYSIS

2.5.1 *Analysis using well-visit + flu comparator*

We repeated the primary analysis, described above, using the well-visit + flu comparator group instead of the well-visit comparator group. The process to determine double negative controls from the list of potential NCEs and NCOs was repeated using the subset of the data which

excluded well-visits among those who did not also have an influenza vaccination in the prior year. Since receipt of influenza vaccine was used to define the well-visit + flu comparator group, 'receipt of influenza vaccine in the prior year' was not used as a measured confounder or as a potential NCE in sensitivity analyses.

2.5.2 *Comparing across double negative controls*

To assess the impact of the choice of one double negative control over another in the primary analysis, estimates and confidence intervals were calculated using the entire set of feasible double negative controls for one non-cardiovascular target outcome (pneumonia) and one cardiovascular target outcome (acute MI). All potential double negative controls that provided sufficient information to adjust for bias in the dataset, as determined by Step 3 in Figure 2, were included in this sensitivity analysis.

Chapter 3. RESULTS

3.1 SELECTING DOUBLE NEGATIVE CONTROLS

One primary double negative control was chosen for each of the four target outcomes using the potential NCEs and NCOs described in Subsection 2.4.2. Diagnosis of lipoma and ingrowing nail in the prior year were present in 0.70% and 0.84% of participants respectively and were therefore eliminated as potential negative controls due to their rarity. After all paired combinations of remaining NCEs and NCOs were assessed for usability, there were 12 potential double negative controls for gout, 3 for pneumonia, 10 for acute MI, and 6 for stroke. Listed in Table 1, the primary double negative controls that were chosen were dermatology visit and hip fracture for gout, prior history of ZVL and hip fracture for both acute MI and pneumonia, and dermatology

visit and epistaxis for stroke. Table 2 details the definitions for all negative control variables that were selected for use in the primary analysis.

Table 2: Operational definitions of negative control variables used in primary analysis.

	Operational Definition [a]	Medical Setting	Pre/Post-vaccination Interval [b]
Negative Control Exposure			
Prior history of ZVL	CVX: 121	IP, ED/UC, OP	Anytime on or prior to index date
Dermatology Visit	Department: DRM	IP, ED/UC, OP	1-365 days prior to index date
Negative Control Outcome			
Hip Fracture	ICD10: S72.0, S72.1, S72.2	IP, ED/UC, OP	1-42 days after index date
Epistaxis	ICD10: R04.0	IP, ED/UC, OP	1-42 days after index date

[a] 4-digit ICD-10 codes included those that started with those 4 digits and had any 5th digit.
[b] Index date refers to the date of the indexed RZV vaccination or well-visit.

3.2 BASELINE CHARACTERISTICS

Table 3 details baseline characteristics of RZV recipients, well-visit comparators, and the well-visit + flu subgroup. A total of 650,701 participants received dose 1 or 2 of RZV in the two-year study period. There were 1,869,067 well-visits indexed in the study period and 1,089,165 well-visits for which the study participant had also received an influenza vaccine in the prior year. RZV recipients were, on average, slightly older than well-visit comparators. They were more likely to seek preventative care such as vaccinations, dermatology visits, and optometry/ophthalmology visits. They were also more likely to have a record of comorbid conditions such as hypertension, hyperlipidemia, and GERD. The distribution of gender was similar across groups, although a slightly higher proportion of women received RZV. The prevalence of many baseline characteristics of the well-visit + flu subgroup fell between the RZV recipients and the well-visit comparators. The proportion of participants from each site in the

well-visit + flu subgroup was approximately the same as in the overall well-visit group, suggesting that influenza vaccination rates were relatively consistent between sites.

Table 3: Characteristics of the study population.

	RZV	Wellness Visit	Wellness + Flu Visit	Total
	N=650701	N=1869067	N=1089165	N=2519768
Age Group, N(%)				
50-54	39785 (6%)	298638 (16%)	124000 (11%)	338423 (13%)
55-59	67529 (10%)	324672 (17%)	148183 (14%)	392201 (16%)
60-64	123925 (19%)	322079 (17%)	171533 (16%)	446004 (18%)
65-69	135332 (21%)	317421 (17%)	203573 (19%)	452753 (18%)
70-74	128188 (20%)	257261 (14%)	180860 (17%)	385449 (15%)
75-79	80775 (12%)	161275 (9%)	118746 (11%)	242050 (10%)
80+	75167 (12%)	187721 (10%)	142270 (13%)	262888 (10%)
Site, N(%)				
1	218063 (34%)	903726 (48%)	524413 (48%)	1121789 (45%)
2	45863 (7%)	149318 (8%)	96260 (9%)	195181 (8%)
3	67344 (10%)	113213 (6%)	60050 (6%)	180557 (7%)
4	11073 (2%)	72225 (4%)	36636 (3%)	83298 (3%)
5	34839 (5%)	60121 (3%)	36348 (3%)	94960 (4%)
6	211922 (33%)	447690 (24%)	273665 (25%)	659612 (26%)
7	61597 (9%)	122774 (7%)	61793 (6%)	184371 (7%)
Female, N(%)	378475 (58%)	1008879 (54%)	600488 (55%)	1387354 (55%)
Healthy user indicator, N(%)				
Prior Zostavax	374431 (58%)	746049 (40%)	563930 (52%)	1120480 (44%)
Influenza Vaccine	531022 (82%)	1089165 (58%)	1089165 (100%)	1620187 (64%)
OPT/OPH Visit	319407 (49%)	671436 (36%)	458493 (42%)	990843 (39%)
Dermatology Visit	150559 (23%)	236787 (13%)	168111 (15%)	387346 (15%)
Pneumonia Vaccine	80648 (12%)	117319 (6%)	91999 (8%)	197967 (8%)
Comorbidity in Year Prior, N (%)				
Diabetes	116932 (18%)	269938 (14%)	186006 (17%)	386870 (15%)
Hypertension	283139 (44%)	585135 (31%)	402577 (37%)	868274 (34%)
Dementia	13328 (2%)	35070 (2%)	25477 (2%)	48398 (2%)
Hyperlipidemia	315577 (48%)	585951 (31%)	411335 (38%)	901528 (36%)
Actinic Keratosis	144356 (22%)	206408 (11%)	149154 (14%)	350764 (14%)
Skin Cancer	21022 (3%)	32490 (2%)	23858 (2%)	53512 (2%)
GERD	130784 (20%)	242394 (13%)	175818 (16%)	373178 (15%)

Osteoarthritis	116579 (18%)	200758 (11%)	142046 (13%)	317337 (13%)
Allergic Rhinitis	39144 (6%)	74711 (4%)	51966 (5%)	113855 (5%)
Ischemic Heart Conditions [a]	58787 (9%)	114372 (6%)	82537 (8%)	173159 (7%)
CHF	19442 (3%)	40638 (2%)	30248 (3%)	60080 (2%)
COPD	28177 (4%)	58577 (3%)	43286 (4%)	86754 (3%)
Metastatic Cancer	12150 (2%)	21702 (1%)	15894 (1%)	33852 (1%)
Chronic Kidney Disease	4952 (1%)	9998 (1%)	7655 (1%)	14950 (1%)
Atrial Fibrillation	37872 (6%)	68699 (4%)	51310 (5%)	106571 (4%)

[a] Presence of ischemic heart disease, transient ischemic attack, or prior stroke.

3.3 COMPARING RELATIVE RISK ESTIMATES FOR SAFETY OUTCOMES ACROSS METHODS

A comparison of RR estimates and confidence intervals for each of the four target safety outcomes is in Table 4. In flexible propensity score-adjusted analyses with the well-visit comparator group, the RR of both acute MI and pneumonia was significantly less than 1, suggesting a potential protective effect of RZV. The RR of gout was significantly greater than 1, suggesting increased risk for RZV recipients. The RR of stroke did not differ significantly from 1 ($p>0.05$). When estimating RR and confidence intervals using the negative control methodology, we found no significant elevated risk or protective effect for RZV recipients for any of the 4 target outcomes ($p>0.05$). In particular, the previously protective effects or RZV that were observed for acute MI and pneumonia attenuated toward 1. Use of negative control variables also resulted in higher RR estimates and larger confidence intervals for acute MI, pneumonia, and gout.

Table 4: Relative risk of target safety outcomes between RZV vaccinees and well-visit comparators.

	Analysis Used	
Safety Outcome	Negative Control Method	Propensity-Adjusted Analysis
Acute MI	1.03 (95% CI: 0.71, 1.48)	0.84 (95% CI: 0.73, 0.95)*
Stroke	0.89 (95% CI: 0.78, 1.02)	0.90 (95% CI: 0.77, 1.02)
Pneumonia	1.05 (95% CI: 0.74, 1.49)	0.89 (95% CI: 0.80, 0.98)*
Gout	1.13 (95% CI: 0.98, 1.30)	1.08 (95% CI: 1.02, 1.14)*

All analyses adjusted for sex, study site, age group (50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+), and select healthy user indicators. Healthy user indicators were influenza vaccine, dermatology visit, optometry/ophthalmology visit, and prior Zostavax vaccine. If an indicator was included as a negative control exposure, that indicator was not used as a covariate. Analyses of cardiovascular outcomes (e.g., acute MI, stroke) additionally adjusted for the cardiovascular risk factors hypertension, hyperlipidemia, diabetes, and ischemic heart condition (presence of ischemic heart disease, TIA, or stroke).
*Indicates the confidence interval does not include 1.

3.1 SENSITIVITY ANALYSIS

3.1.1 *Analysis using well-visit + flu comparator*

Repeating the process of determining double negative controls in the subset composed of well-visit + flu and RZV recipients resulted in different double negative controls for gout and acute MI but not for pneumonia or stroke. The double negative control used for RR estimates of gout and pneumonia was prior receipt of ZVL and hip fracture. The double negative control used for RR estimates of stroke and acute MI was dermatology visit and epistaxis. Table 5 presents the well-visit + flu comparator results of negative control analyses versus the original analysis. As hypothesized, estimates and confidence intervals based on the well-visit + flu subset comparators calculated with and without negative controls were extremely similar for all target outcomes, with the exception of pneumonia. The RR of pneumonia calculated without double negative control was less than 1 and statistically significant (RR: 0.87; 95% CI: 0.78, 0.96), suggesting a protective effect. In contrast, the RR of pneumonia calculated with double negative control was

greater than 1 but not statistically significant (RR: 1.03; 95% CI: 0.74, 1.42). Without the use of double negative controls, the conclusions drawn from analysis using the well-visit + flu comparator group (i.e.., protective effect for acute MI and for pneumonia, null effect for stroke, and increased risk of gout) did not differ from the conclusions drawn from analysis using the primary well-visit comparator group, although estimates were slightly closer to 1.

Table 5: Relative risk of target safety outcomes between RZV vaccinees and well-visit + flu comparators.

	Analysis Used	
Safety Outcome	Negative Control Method	Propensity-Adjusted Analysis
Acute MI	0.85 (95% CI: 0.73, 0.95) *	0.85 (95% CI: 0.73, 0.96) *
Stroke	0.92 (95% CI: 0.79, 1.07)	0.94 (95% CI: 0.80, 1.08)
Pneumonia	1.03 (95% CI: 0.74, 1.42)	0.87 (95% CI: 0.78, 0.96) *
Gout	1.07 (95% CI: 1.00, 1.16)	1.06 (95% CI: 1.00, 1.13) *

All analyses adjusted for sex, study site, age group (50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+), and select healthy user indicators. Healthy user indicators were dermatology visit, optometry/ophthalmology visit, and prior Zostavax vaccine. If an indicator was included in the analysis as a negative control exposure that indicator was not used as a covariate. Analyses of cardiovascular outcomes (e.g., acute MI, stroke) additionally adjusted for the cardiovascular risk factors hypertension, hyperlipidemia, diabetes, and ischemic heart condition (presence of ischemic heart disease, TIA, or stroke).
* Indicates that confidence interval does not include 1

3.1.2 *Comparing estimates across double negative controls*

Conclusions drawn from the RR estimates and confidence intervals differed somewhat depending on which double negative control was used in the analysis. Estimates of the RR of acute MI when using the well visit comparator group ranged from 0.78 to 1.03 across the ten double negative controls that were determined to provide sufficient information to correct for unmeasured confounding (Figure 3). Seven double negative controls suggested a protective effect for RZV recipients while the other three resulted in RR estimates that did not differ significantly from 1 ($p>0.05$), although confidence intervals were all overlapping. Estimates and confidence intervals calculated for the RR of pneumonia using different double negative controls

also exhibited some variability (see Appendix A, Table 8). Out of the three double negative controls that were determined to provide sufficient information for pneumonia, one suggested a protective effect for RZV recipients (RR: 0.89; 95% CI: 0.81, 0.98), while the other two resulted in RR estimates which did not differ significantly from 1 (RR: 1.05; 95% CI: 0.74, 1.49; and RR: 0.96; 95% CI: 0.81, 1.13).

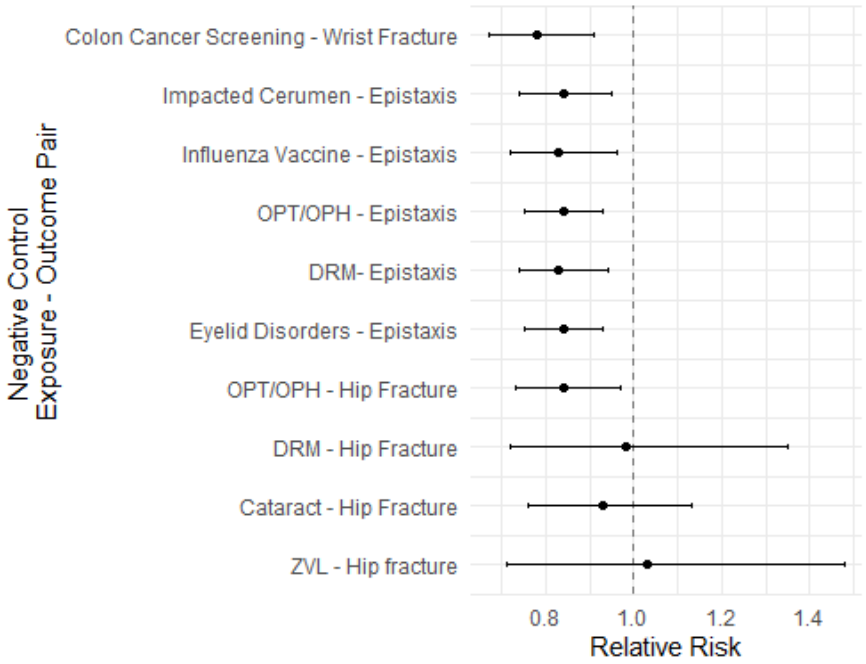

Figure 3: Relative risk of acute myocardial infarction between RZV vaccinees and well-visit comparators estimated using different double negative controls.

Chapter 4. DISCUSSION

Confounding bias is a key concern in the electronic health record setting in which data have been gathered for the purposes of clinical care rather than research and thus not all confounders are proactively measured. We were able to successfully leverage negative control variables to demonstrate how correction for unmeasured confounding can be conducted in such a setting by developing a principled approach to selection of double negative controls and by implementing,

for the first time in practice, a multiply robust causal inference method with double negative controls. Using double negative control methodology led to slightly different conclusions than the more traditional confounder adjustment approach. When conclusions differed, they differed in the hypothesized direction; protective effects that were observed using flexible propensity score regression attenuated and were no longer statistically significant when double negative controls were applied. This suggests that the protective effects observed using standard regression adjustment methods may have been due to healthy user bias instead of receipt of RZV. The increased risk of gout, which was observed in the flexible propensity-adjusted analysis, was also not statistically significant when estimated using double negative controls. This is in part explained by the widening of confidence intervals, which was often observed when using double negative controls, and likely reflects a combination of the rarity of the negative controls used and an appropriate acknowledgment of additional uncertainty in the RR estimate due to residual systematic error that traditional methods may not take into account.

We hypothesized that the results of the analyses with and without double negative controls would be more similar when using the well-visit + flu comparator group than when using the well-visit comparator group. The reasoning behind this hypothesis was that well-visit + flu comparators exhibited higher proportions of health-seeking behavior than well-visit comparators in the original RZV safety study, thus reducing the amount of bias eliminated by double negative control methodology by reducing the amount of potential healthy user bias to eliminate. Our results supported this hypothesis; estimates and confidence intervals were more consistent between the double negative control analysis and the flexible propensity score regression analysis when well-visit + flu was used as the comparator than when well-visit was used as the comparator. Like RZV recipients, the well-visit + flu subset exhibited more health-

seeking behavior than the well-visit group, which would have reduced the potential for healthy user bias in the sensitivity analysis. This supports the idea that double negative control methodology was able to correct for healthy user bias in the primary analysis.

A major advantage of using double negative controls is that it allows us to adjust for unmeasured confounding between two independent groups, making the use of concurrent comparators a much more viable option. In vaccine safety surveillance studies, vaccinated and unvaccinated individuals are often fundamentally different, necessitating the use of self-controlled design or historical comparators, which have important known weaknesses.[60-66] Limitations of self-controlled study designs include susceptibility to bias when event rates are not constant within the study period.[65] When using historical comparators or self-controlled design, temporal trends in the target outcome, or in diagnostic or coding practices may lead to false signaling or failure to identify a true signal.[60,65] Using a contemporary comparator group avoids risk of temporal bias.

Another advantage of double negative control methods is the relaxation of methodological assumptions. For instance, perfect measurement of covariates and lack of unmeasured confounding are often required in other causal inference methods in order to fully adjust for confounding but are not necessary when using double negative control methodology. In addition, double negative control methods do not require assumptions which are often necessary for other methods that allow for unmeasured confounding or imperfect measurement of covariates. Such assumptions include monotonicity, rank preservation, and the assumption of a linear model for unmeasured confounding.[24,37-39] The identification assumptions required for double negative control methods are outlined in Appendix C. Typical to causal inference, double negative control methods assume positivity and consistency. The consistency assumption cannot

be verified empirically, relying instead on a well-defined treatment of interest. The plausibility of the positivity assumption can be verified empirically by checking whether each treatment group for both the NCE and the treatment of interest contain participants with every possible combination of measured covariates. Double negative control methods assume latent ignorability, relaxing the more common ignorability assumption to allow for the existence of unmeasured confounders.[41] The use of double negative control methods may therefore be preferable in situations where we cannot plausibly assume that there are no unmeasured confounders. Assumptions 3, 4, 5, and 7 in the appendix detail conditions of the negative controls, and their relationships to the unmeasured confounder. These assumptions rely on scientific and subject matter knowledge and cannot be verified empirically. A study should use double negative control methods to control for bias only if that study has the capacity to collect NCEs and NCOs that plausibly meet these assumptions. If the key assumptions outlined in Appendix C are not plausible in the context of a given study, then double negative control methods should not be used.

In this paper we have outlined a principled approach which can be used to choose a primary double negative control for a given outcome and treatment of interest. One current weakness in the use of double negative controls to adjust for unmeasured confounders is the lack of underlying theory behind an ideal choice of double negative control. The sensitivity analysis that we performed using multiple different negative controls to estimate a particular vaccine-outcome association demonstrates that the choice of double negative control has the potential to influence the conclusions drawn from the analysis. As seen in Figure 3, two double negative controls with the same NCE but different NCOs would sometimes, but not always, result in different estimates and confidence intervals for the same target outcome. Double negative

controls with the same NCO but different NCEs would often, but not always, result in similar estimates and confidence intervals for the same target outcome. However, it is worth noting that there was no reversal of an observed protective effect to an observed increased risk or vice versa; differences in conclusions were limited to the difference between a protective effect and an effect that was not statistically significant. Since this study used a real-world dataset, we do not know the objective truth of the causal effect, and so cannot know which double negative control resulted in an estimate closest to 'truth.' Future research that simulates negative control variables with a variety of different characteristics would allow us to more fully understand the observable properties of a double negative control that result in the most accurate estimates while still acknowledging the potential uncertainty due to unmeasured confounding.

This study relied on electronic medical data collected during the course of routine health care, not for research purposes. The use of double negative control methods helps deal with unmeasured or mismeasured confounders, a key limitation of such data. However, it does not compensate for all limitations of the data. Limitations which remain include incomplete data capture due to delays in receipt of some insurance claims by MCO systems, misclassification of adverse events, and misclassification of vaccination status.[60-65] In particular, the health outcomes that were captured in this dataset do not necessarily reflect the true health condition status present in each participant. A health condition was recorded only if the participant received a diagnosis, meaning that a health condition would go unrecorded if a participant did not seek medical care. If we assume that RZV recipients are more likely to seek preventative care, it follows that RZV recipients would be more likely to seek medical care for a health condition and would therefore be more likely to be diagnosed with existing conditions. Such an effect would not extend to critical conditions for which the majority of individuals would seek medical care,

regardless of their typical health-seeking behavior. This may help explain why RZV recipients demonstrate higher rates of comorbid conditions such as hypertension, hyperlipidemia, and gastroesophageal reflux disease (GERD). This may also help explain why the increased risk of gout observed in analyses without double negative control methods was no longer statistically significant when double negative control methods were used; participants exhibiting more health-seeking behavior may have been more likely to seek medical aid for gout.

Double negative control methods often necessitate the collection of additional measures to use as negative control variables. Identifying and selecting potential negative control variables requires making assumptions about causal relationships (or lack thereof) between measures. For this reason, drug product labels and medical literature can be invaluable resources when choosing potential negative control variables. However, causal relationships between primary variables and potential negative controls are often difficult, if not impossible, to verify. Additionally, collecting extra measures to use as negative controls may add time or monetary costs to a study. Proximal causal inference, a generalized framework of double negative control methods, also uses two independent proxy variables to nonparametrically identify causal effect. But it lends the added flexibility of allowing the use of measures that would typically be gathered to adjust for confounders as independent proxy variables instead of exclusively relying on negative control variables. This reduces the need to collect excess variables in order to identify the confounding bridge necessary to correct for unmeasured confounding.[45,46]

Double negative control methods are a promising new way to correct for healthy user bias without making restrictive modeling assumptions. Future research will simulate negative control variables in order to better define the observable properties of an ideal double negative control in real-world data. Another direction of further research is the continued extension of the

theoretical framework of double negative control methods to the more general proximal causal inference. A future study will apply this more general method in a secondary analysis of a pharmacosurveillance study using independent proxy variables that were originally gathered for use as covariates in a more traditional analysis such as outcome regression, IPTW, matching or stratification.

REFERENCES

1. Matthews KA, Kuller LH, Wing RR, Meilahn EN, Plantinga P. Prior to Use of Estrogen Replacement Therapy, Are Users Healthier than Nonusers? *American Journal of Epidemiology*. 1996;143(10):971-978. doi:10.1093/oxfordjournals.aje.a008678

2. Garbe E, Suissa S. Issues to debate on the Women's Health Initiative (WHI) study: Hormone replacement therapy and acute coronary outcomes: methodological issues between randomized and observational studies. *Human Reproduction*. 2004;19(1):8-13. https://doi.org/10.1093/humrep/deh022

3. Zhang X, Faries DE, Li H, Stamey JD, Imbens GW. Addressing unmeasured confounding in comparative observational research. *Pharmacoepidemiology and Drug Safety*. 2018;27(4):373-382. doi:10.1002/pds.4394

4. Kim M-C, Yun S-C, Lee S-O, et al. Statins increase the risk of herpes zoster: A propensity score-matched analysis. Wright JM, ed. *PLOS ONE*. 2018;13(6):e0198263. doi:10.1371/journal.pone.0198263

5. Rasmussen JN, Chong A, Alter DA. Relationship Between Adherence to Evidence-Based Pharmacotherapy and Long-term Mortality After Acute Myocardial Infarction. *JAMA*. 2007;297(2):177. doi:10.1001/jama.297.2.177

6. Setoguchi S, Glynn RJ, Avorn J, Mogun H, Schneeweiss S. Statins and the Risk of Lung, Breast, and Colorectal Cancer in the Elderly. *Circulation*. 2007;115(1):27-33. doi:10.1161/circulationaha.106.650176

7. Shrank WH, Patrick AR, Alan Brookhart M. Healthy User and Related Biases in Observational Studies of Preventive Interventions: A Primer for Physicians. *Journal of General Internal Medicine*. 2011;26(5):546-550. doi:10.1007/s11606-010-1609-1

8. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011

9. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Statistical Methods in Medical Research*. 2009;18(1):67-80. doi:10.1177/0962280208092347

10. Xu S, Shetterly S, Cook AJ, et al. Evaluation of propensity scores, disease risk scores, and regression in confounder adjustment for the safety of emerging treatment with group sequential monitoring. *Pharmacoepidemiology and Drug Safety*. 2016;25(4):453-461. doi:10.1002/pds.3983

11. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment

effects in observational studies. *Statistics in Medicine*. 2015;34(28):3661-3679. doi:10.1002/sim.6607

12. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Medical Care*. 2007;45(10 Supl 2):S143-S148. doi:10.1097/MLR.0b013e318074ce79

13. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512-522. doi:10.1097/EDE.0b013e3181a663cc

14. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *International Journal of Epidemiology*. 2005;35(2):337-344. doi:10.1093/ije/dyi274

15. Simonsen L, Reichert TA, Viboud C, Blackwelder WC, Taylor RJ, Miller MA. Impact of influenza vaccination on seasonal mortality in the US elderly population. *Archives of Internal Medicine*. 2005;165(3):265–72. doi:10.1001/archinte.165.3.265

16. Vu T, Farish S, Jenkins M, Kelly H. A meta-analysis of effectiveness of influenza vaccine in persons aged 65 years and over living in the community. *Vaccine*. 2002;20(13-14):1831-1836. doi:10.1016/s0264-410x(02)00041-5

17. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating Short-Term Drug Effects Using a Physician-Specific Prescribing Preference as an Instrumental Variable. *Epidemiology*. 2006;17(3):268-275. doi:10.1097/01.ede.0000193606.58671.c5

18. Brookhart MA, Schneeweiss S. Preference-Based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results. *The International Journal of Biostatistics*. 2007;3(1). doi:10.2202/1557-4679.1072

19. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 91(434):444–472. doi: 10.1080/01621459.1996.10476902

20. Terza JV, Bradford WD, Dismuke CE. The Use of Linear Instrumental Variables Methods in Health Services Research and Health Economics: A Cautionary Note. *Health Services Research*. 2007;43(3):1102-1120. doi:10.1111/j.1475-6773.2007.00807.x

21. Staiger D, Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica*. 1997;65(3):557. doi:10.2307/2171753

22. Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in Medicine*. 2007;27(9):1539-1556. doi:10.1002/sim.3036

23. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*. 2000;29(4):722-729. doi:10.1093/ije/29.4.722

24. Swanson SA, Miller M, Robins JM, Hernán MA. Definition and Evaluation of the Monotonicity Condition for Preference-based Instruments. *Epidemiology*. 2015;26(3):414-420. doi:10.1097/ede.0000000000000279

25. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream?. *Epidemiology*. 2006;17(4):360-372. doi:10.1097/01.ede.0000222409.00878.37

26. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb

27. Lousdal ML, Lash TL, Flanders WD, et al. Negative controls to detect uncontrolled confounding in observational studies of mammographic screening comparing participants and non-participants. *International Journal of Epidemiology*. 2020;49(3). doi:10.1093/ije/dyaa029

28. Yerushalmy J. The relationship of parents' cigarette smoking to outcome of pregnancy—implications as to the problem of inferring causation from observed associations1. *International Journal of Epidemiology*. 2014;43(5):1355-1366. doi:10.1093/ije/dyu160

29. Smith GD. Assessing Intrauterine Influences on Offspring Health Outcomes: Can Epidemiological Studies Yield Robust Findings? *Basic & Clinical Pharmacology & Toxicology*. 2008;102(2):245-256. doi:10.1111/j.1742-7843.2007.00191.x

30. Trichopoulos D, Zavitsanos X, Katsouyanni K, Tzonou A, Dalla-Vorgia P. Psychological stress and fatal heart attack: the Athens (1981) earthquake natural experiment. *The Lancet*. 1983;321(8322):441-444. doi:10.1016/s0140-6736(83)91439-3

31. Sheppard L, Levy D, Norris G, Larson TV, Koenig JQ. Effects of Ambient Air Pollution on Nonelderly Asthma Hospital Admissions in Seattle, Washington, 1987–1994. *Epidemiology*. 1999;10(1):23-30. doi:10.1097/00001648-199901000-00006

32. Norén GN, Bergvall T, Ryan PB, Juhlin K, Schuemie MJ, Madigan D. Empirical Performance of the Calibrated Self-Controlled Cohort Analysis Within Temporal Pattern Discovery: Lessons for Developing a Risk Identification and Analysis System. *Drug Safety*. 2013;36(S1):107-121. doi:10.1007/s40264-013-0095-x

33. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p -values. *Statistics in Medicine*. 2013;33(2):209-218. doi:10.1002/sim.5925

34. Richardson DB, Laurier D, Schubauer-Berigan MK, Tchetgen ET, Cole SR. Assessment and indirect adjustment for confounding by smoking in cohort studies using relative

hazards models. *American Journal of Epidemiology.* 2014;180(9):933-40. doi:10.1093/aje/kwu211

35. Sanderson E, Macdonald-Wallis C, Smith GD, Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration, *International Journal of Epidemiology*. 2018; 47(2):597-596. https://doi.org/10.1093/ije/dyx213

36. Glynn A, Ichino N. Generalized Nonlinear Difference-in-Difference-in-Differences (June 1, 2019). *V-Dem Working Paper*. 2019:90. http://dx.doi.org/10.2139/ssrn.3410888

37. Tchetgen Tchetgen EJ, Sofer T, Richardson D. Negative outcome control for unobserved confounding under a Cox proportional hazards model. *Harvard University Biostatistics Working Paper* Series. 2015; Working Paper 192. https://biostats.bepress.com/harvardbiostat/paper192

38. Tchetgen Tchetgen E. The Control Outcome Calibration Approach for Causal Inference With Unobserved Confounding. *American Journal of Epidemiology*. 2014;179(5):633-640. doi:10.1093/aje/kwt303

39. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics*. 2011;13(3):539-552. doi:10.1093/biostatistics/kxr034

40. Sofer T, Richardson DB, Colicino E, Schwartz J, Tchetgen Tchetgen EJ. On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences. *Statistical Science.* 2016;31(3):348-361. doi:10.1214/16-sts558

41. Miao W, Geng Z, Tchetgen Tchetgen EJ. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika.* 2018;105(4):987–993. doi:10.1093/biomet/asy038

42. Shi X, Miao W, Tchetgen Tchetgen ET. A Selective Review of Negative Control Methods in Epidemiology. *Current Epidemiology Reports*. 2020;7(4):190-202. doi:10.1007/s40471-020-00243-4

43. Shi X, Miao W, Nelson J, Tchetgen Tchetgen EJ. Multiply robust causal inference with double negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2020;82(2):521–540. doi:10.1111/rssb.12361

44. Shi, X., Miao, W. & Tchetgen Tchetgen, EJ. A Confounding Bridge Approach for Double Negative Control Inference on Causal Effects. 2020. In progress, a prior version can be found at https://arxiv.org/abs/1808.04945.

45. Tchetgen Tchetgen EJ, Ying A, Cui Y, Shi X, Miao W. An Introduction to Proximal Causal Learning. 2020, doi:10.1101/2020.09.21.20198762.

46. Cui Y, Pu H, Shi X, Miao W. Tchetgen Tchetgen EJ, Ying. Semiparametric Proximal Causal Inference. 2020, doi: https://doi.org/10.1101/2020.09.21.20198762

47. Dooling KL, Guo A, Patel M, et al. Recommendations of the Advisory Committee on Immunization Practices for Use of Herpes Zoster Vaccines. *MMWR Morbidity and Mortality Weekly Report*. 2018;67(3):103-108. doi:10.15585/mmwr.mm6703a5

48. Food and Drug Administration. Shingrix [package insert]. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration; 2017. https://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM581605.pdf

49. Brookhart MA, Patrick AR, Dormuth C, et al. Adherence to Lipid-lowering Therapy and the Use of Preventive Health Services: An Investigation of the Healthy User Effect. *American Journal of Epidemiology*. 2007;166(3):348-354. doi:10.1093/aje/kwm070

50. Nelson JC, Ulloa-Pérez E, Yu O, et al. Active Post-Licensure Safety Surveillance for Recombinant Zoster Vaccine [Unpublished manuscript]. 2021

51. López-Fauqued M, Campora L, Delannois F, et al. Safety profile of the adjuvanted recombinant zoster vaccine: Pooled analysis of two large randomised phase 3 trials. *Vaccine*. 2019;37(18). doi:10.1016/j.vaccine.2019.03.043

52. Shi X, Wellman R, Heagerty PJ, Nelson JC, Cook AJ. Safety surveillance and the estimation of risk in select populations: Flexible methods to control for confounding while targeting marginal comparisons via standardization. *Statistics in Medicine*. 2020;39(4):369-386. doi:10.1002/sim.8410

53. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

54. Stefanski LA, Boos DD. The Calculus of M-Estimation. *The American Statistician*. 2002;56(1):29-38. doi:10.1198/000313002753631330

55. Dormuth CR, Patrick AR, Shrank WH, et al. Statin adherence and risk of accidents a cautionary tale. *Circulation.* 2009;119(15):2051-2057. doi:10.1161/CIRCULATIONAHA.108.824151

56. Tseng HF. Herpes Zoster Vaccine in Older Adults and the Risk of Subsequent Herpes Zoster Disease. *JAMA*. 2011;305(2):160. doi:10.1001/jama.2010.1983

57. Izurieta HS, Wernecke M, Kelman J, et al. Effectiveness and Duration of Protection Provided by the Live-attenuated Herpes Zoster Vaccine in the Medicare Population Ages 65 Years and Older. *Clinical Infectious Diseases*. 2017;64(6):785-793. doi:10.1093/cid/ciw854

58. Leung J, Harpaz R, Molinari N-A, Jumaan A, Zhou F. Herpes Zoster Incidence Among Insured Persons in the United States, 1993–2006: Evaluation of Impact of Varicella Vaccination. *Clinical Infectious Diseases*. 2011;52(3):332-340. doi:10.1093/cid/ciq077

59. Setoguchi S, Warner Stevenson L, Stewart GC, et al. Influence of healthy candidate bias in assessing clinical effectiveness for implantable cardioverter-defibrillators: cohort study of older patients with heart failure. *BMJ*. 2014;348(may08 3):g2866-g2866. doi:10.1136/bmj.g2866

60. Lieu TA, Kulldorff M, Davis RL, et al. Real-Time Vaccine Safety Surveillance for the Early Detection of Adverse Events. *Medical Care*. 2007;45(10):S89-S95. doi:10.1097/mlr.0b013e3180616c0a

61. Klein NP, Fireman B, Yih WK, et al. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics*. 2010;126(1):e1-8. doi:10.1542/peds.2010-0665

62. Yih WK, Nordin JD, Kulldorff M, et al. An assessment of the safety of adolescent and adult tetanus–diphtheria–acellular pertussis (Tdap) vaccine, using active surveillance for adverse events in the Vaccine Safety Datalink. *Vaccine*. 2009;27(32):4257-4262. doi:10.1016/j.vaccine.2009.05.036

63. Belongia EA, Irving SA, Shui IM, et al. Real-Time Surveillance to Assess Risk of Intussusception and Other Adverse Events After Pentavalent, Bovine-Derived Rotavirus Vaccine. *Pediatric Infectious Disease Journal*. 2010;29(1):1-5. doi:10.1097/inf.0b013e3181af8605

64. Gee J, Naleway A, Shui I, et al. Monitoring the safety of quadrivalent human papillomavirus vaccine: Findings from the Vaccine Safety Datalink. *Vaccine*. 2011;29(46):8279-8284. doi:10.1016/j.vaccine.2011.08.106

65. Greene SK, Kulldorff M, Yin R, et al. Near real-time vaccine safety surveillance with partially accrued data. *Pharmacoepidemiology and Drug Safety*. 2011;20(6):583-590. doi:10.1002/pds.2133

66. Whitaker HJ, Paddy Farrington C, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Statistics in Medicine*. 2006;25(10):1768-1797. doi:10.1002/sim.2302

67. Rubin DB, Schenker N. Multiple imputation in health-are databases: An overview and some applications. *Statistics in Medicine*. 1991;10(4):585-598. doi:10.1002/sim.4780100410

68. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting Effect Estimates for Unmeasured Confounding with Validation Data using Propensity Score Calibration. *American Journal of Epidemiology*. 2005;162(3):279-289. doi:10.1093/aje/kwi192

69. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *Journal of clinical epidemiology*. 2015;68(2):122-133. doi:10.1016/j.jclinepi.2014.06.021

70. Ji X, Small DS, Leonard CE, Hennessy S. The Trend-in-trend Research Design for Causal Inference. *Epidemiology*. 2017;28(4):529-536. doi:10.1097/ede.0000000000000579

APPENDIX A: SUPPLEMENTARY TABLES

Table 6: Operational definitions of covariates used in analysis.

Covariate	Operational Definitions [a]	Medical Setting	Pre-vaccination Interval
Hyperlipidemia	ICD-10: E78	IP, ED/UC, OP	1-365 days prior to index date
Hypertension	ICD-10: I10	IP, ED/UC, OP	1-365 days prior to index date
Diabetes	ICD-10: E08, E10, E11	IP, ED/UC, OP	1-365 days prior to index date
Ischemic Heart Condition	ICD-10: I20, I21, I22, I23, I24, I25, I63, I61.9, G45	IP, ED/UC, OP	1-365 days prior to index date
Receipt of Influenza Vaccine	CVX: 151, 135, 153, 171, 186, 158, 150, 161, 166, 111, 149, 155, 185, 141, 140, 144, 15, 168, 88, 16	IP, ED/UC, OP	1-365 days prior to index date
Receipt of ZVL	CVX: 121	IP, ED/UC, OP	
Optometry/ Ophthalmology Visit	Department: OPH, OPT	IP, ED/UC, OP	1-365 days prior to index date
Dermatology Visit	Department: DRM	IP, ED/UC, OP	1-365 days prior to index date

[a] 3-digit ICD-10 codes (e.g., 345) included those that started with those 3 digits and contained any additional 4th or 5th digits (e.g., 345.11); 4-digit ICD-10 codes included those that started with those 4 digits and had any 5th digit.

Table 7: Operational definitions of all negative control variables considered for potential use in analysis.

Negative Control Variable	Operational Definitions	Medical Setting	Pre/Post-vaccination Interval (days)[b]
	Exposure		
Influenza Vaccine	CVX: 151, 135, 153, 171, 186, 158, 150, 161, 166, 111, 149, 155, 185, 141, 140, 144, 15, 168, 88, 16	IP, ED/UC, OP	1-365 days prior to index date
Zostavax	CVX: 121	IP, ED/UC, OP	1-365 days prior to index date

OPT/OPH Visit	Department: OPH, OPT	IP, ED/UC, OP	1-365 days prior to index date
Dermatology Visit	Department: DRM	IP, ED/UC, OP	1-365 days prior to index date
Glaucoma	ICD10: H40	Outpatient	1-365 days prior to index date
Colon Cancer Screening	ICD10: Z12.11	Outpatient	1-365 days prior to index date
Impacted Cerumen	ICD10: H61.2	Outpatient	1-365 days prior to index date
Eyelid Disorders	ICD10: H00, H01.0	Outpatient	1-365 days prior to index date
Cataract	ICD10: H25	Outpatient	1-365 days prior to index date
Lipoma	ICD10: D17	Outpatient	1-365 days prior to index date
Ingrowing nail	ICD10: L60.0	Outpatient	1-365 days prior to index date
	Outcome		
Hip fracture	ICD10: S72.0, S72.1, S72.2	IP, ED/UC, OP	1-42 days after index date
Wrist fracture	ICD10: S62.0, S52.1	IP, ED/UC, OP	1-42 days after index date
Epistaxis	ICD10: R04.0	IP, ED/UC, OP	1-42 days after index date
Leg Cellulitis	ICD10: L031.15, L03.116	IP, ED/UC, OP	1-42 days after index date

[a] 3-digit ICD-10 codes (e.g., 345) included those that started with those 3 digits and contained any additional 4th or 5th digits (e.g., 345.11); 4-digit ICD-10 codes included those that started with those 4 digits and had any 5th digit.
[b] Index date refers to the date of the indexed RZV vaccination or well-visit.

Table 8: Relative risk of target safety outcomes RZV vaccinees and well-visit comparators estimated using different double negative controls

Negative Control Outcome and Exposure	Relative Risk (95% CI)
Pneumonia	
NCE: Prior receipt of ZVL NCO: Hip fracture	1.05 (0.74, 1.49)
NCE: OPT/OPH visit NCO: Epistaxis	0.89 (0.81, 0.98)
NCE: Cataract NCO: Epistaxis	0.96 (0.81, 1.13)
Acute MI	
NCE: Prior receipt of ZVL NCO: Hip fracture	1.03 (0.71, 1.48)
NCE: Cataract NCO: Hip fracture	0.93 (0.76, 1.13)
NCE: Dermatology visit NCO: Hip fracture	0.98 (0.72, 1.35)
NCE: OPT/OPH visit NCO: Hip fracture	0.84 (0.73, 0.97)
NCE: Eyelid disorders NCO: Epistaxis	0.84 (0.75, 0.93)
NCE: Dermatology visit NCO: Epistaxis	0.83 (0.74, 0.94)
NCE: OPT/OPH visit NCO: Epistaxis	0.84 (0.75, 0.93)
NCE: Influenza vaccine NCO: Epistaxis	0.83 (0.72, 0.96)
NCE: Impacted cerumen NCO: Epistaxis	0.84 (0.74, 0.95)
NCE: Colon cancer screening NCO: Wrist fracture	0.78 (0.67, 0.91)

All analyses adjusted for sex, study site, age group (50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+), and select healthy user indicators. Healthy user indicators were influenza vaccine, dermatology visit, optometry/ophthalmology visit, and prior Zostavax vaccine. If an indicator was included in the analysis as a negative control exposure, that indicator was not used as a covariate. Analyses of cardiovascular outcome (acute MI) additionally adjusted for the cardiovascular risk factors hypertension, hyperlipidemia, diabetes, and ischemic heart condition (presence of ischemic heart disease, TIA, or stroke).

APPENDIX B: REVIEW OF SELECT METHODS TO ACCOUNT FOR UNMEASURED CONFOUNDING

There are a wide range of statistical methods which have been developed to account for unmeasured confounding. Each method has its own benefits and drawbacks. This appendix describes a select few. For a more comprehensive summary of possible methods, we recommend review papers such as Zhang et al., 2018.[3]

Some methods, such as multiple imputation and propensity score calibration, can avoid measuring proxies in the main analysis data while retaining unbiased results by using the missing data perspective. Multiple imputation can be used to correct for unmeasured confounding by imputing values of unmeasured confounders based on a subsample of the population in which information on unmeasured confounders was obtained.[67] The adjusted treatment effect can then be estimated using imputed values. Propensity score calibration involves calculating two propensity scores, one incorporating only the measured confounders on the full cohort of interest, and another which incorporates measured and unmeasured confounders (collected from a subsample or from external data). Regressing the first propensity score on the second results in estimated regression coefficients which can be used to calibrate naive and biased results based only on measured confounders in the full cohort.[68] Addressing unmeasured confounding from a missing data perspective is not always a feasible option, as it implies that the unmeasured confounder was measured in some capacity, often in a subsample or from an external source.

Other methods, such as regression discontinuity and trend-in-trend analysis, allow for unbiased estimates in the presence of unmeasured confounding by relying on specific assumptions about the unmeasured confounding variables. Regression discontinuity accounts for unobserved differences between treatment and comparator groups by controlling for the variable

which determines whether an individual will receive treatment. This method relaxes the assumption of no unmeasured confounding by instead relying on the assumption that there is no unmeasured confounding at the treatment threshold; in the small area around the cutoff score which determines treatment assignment, the only difference between treatment and control groups is treatment assignment.[69] Regression discontinuity is specific to the situation in which the assignment of the treatment of interest is based on whether a patient scores above or below a cutoff point on a continuously measured variable. In trend-in-trend analyses, predicted probability of exposure over the entire study period is estimated, then the population is stratified according to the quintile of estimated probability of exposure.[3,70] Treatment effect can be estimated by examining changes in outcome occurrence as a function of changes in exposure prevalence across strata, with differential time trends in exposure. Trend-in-trend analysis is feasible only when there is a strong time trend in the exposure of interest. It allows for unmeasured confounding under specific circumstances; instead of assuming no unmeasured confounding, this method assumes that any time trends in an unmeasured confounder will be uncorrelated with time trends in the exposure of interest across strata.[70] For example, estimates calculated using trend-in-trend analysis would be biased due to unmeasured confounding if the trend in use of a co-intervention was positively correlated with the trend of use of the exposure of interest.

APPENDIX C: SUMMARY OF KEY ASSUMPTIONS FOR NONPARAMETRIC IDENTIFICATION IN A DOUBLE NEGATIVE CONTROL DESIGN[41,42,52]

Assumption 1 (Consistency): $Y(a)=Y$ when $A=a$. The intervention does not change how the exposure may affect the outcome. If researchers assign a given treatment group $A=a$, the outcome $Y(a)$ will be the same as if the treatment group is assigned by nature.

Assumption 2 (Latent ignorability): $A \perp\!\!\!\perp Y(a) \mid U, X$. Causal inference in observational studies often relies on the assumption of ignorability, which states that $A \perp\!\!\!\perp Y(a) \mid X$, where X describes the measured covariates. In a properly randomized clinical trial, the treatment allocation mechanism is independent of the counterfactual outcomes. However, this is often not the case in an observational study. For example, if the treatment is allocated more frequently to early-stage cancer patients than later-stage cancer patients, the treatment allocation mechanism would no longer be independent of counterfactual survival. The ignorability assumption does not allow for unmeasured common causes (U). Here, we rely on the assumption of latent ignorability, instead of the assumption of ignorability to establish causation, allowing for the existence of unmeasured common causes.

Assumption 3 (Negative control outcome): $W(a, z) = W$ and $W \perp\!\!\!\perp A \mid U, X$.

a. When unmeasured and measured confounders (*U*, *X*) are accounted for, there is no remaining common cause between the negative control outcome (*W*) and the target exposure (*A*).
b. There is no causal effect of the exposure of interest and negative control exposure (*Z*) on the negative control outcome given the unmeasured and measured confounders.
c. An exposure (e.g., drug treatment) may be used as a negative control outcome so long as it satisfies the above criteria.

Assumption 4 (Negative control exposure): $Y(a, z) = Y(a)$ and $Z \perp\!\!\!\perp (Y(a), W) \mid U, X$.

a. When unmeasured and measured confounders (*U*, *X*) are accounted for, there is no remaining common cause between the negative control exposure (*Z*) and the target outcome (*Y*) or between the negative control exposure and the negative control outcome (*W*).
b. There is no causal effect of the negative control exposure on the target outcome given the target exposure and the unmeasured and measured confounders.
c. An outcome may be used as a negative control exposure so long as it satisfies the above criteria.

Assumption 5 (U-comparable)[a]: $W \not\perp\!\!\!\perp U \mid X$ and $Z \not\perp\!\!\!\perp U \mid A, X$.

a. Unmeasured confounders of the *A*-*Y* association are identical to those of the *A*-*W* association and Z-Y association; a non-zero *A*-*W* or *Z*-*Y* association can be attributed to *U*.

b. Presence of an association between primary and negative control variables (*A*-*W* or *Z*-*Y*) implies residual confounding bias, while absence of such associations implies no empirical evidence of unmeasured confounding. When evaluating the *Z*-*Y* association, it is necessary to adjust for *A* because of the possibility of $Z \rightarrow A \rightarrow Y$ or $Z \leftarrow A \rightarrow Y$.

Assumption 6 (Positivity): all participants could potentially be assigned to each treatment group. This assumption would not hold true if, for example, there was a high risk and a low risk treatment, and no participant with mild symptoms would ever be assigned to the high risk treatment. $0 < \Pr(A=a, Z=z \mid X) < 1$ for all a, z.

Assumption 7 (Completeness)[a]:

a. For all a, $W \perp\!\!\!\perp Z \mid A = a, X$. Both *Z* and *W* should be associated with *U* such that variation in *U* can be recovered from variation in *Z* and *W*.
b. For any square integrable function g, if $\mathrm{E}[g(W)|Z = z, A = a, X] = 0$ for almost all z, a, then $g(W) = 0$. This assumption aims to ensure that the underlying unmeasured confounding mechanism in $\mathrm{E}[Y \mid A, U]$ can be identified using *Z* and *W*. For example, suppose *U* is a binary variable. Then Assumption 7 further requires that Z and W have at least two categories, and $\mathrm{E}[W \mid A = a, Z = 1, X = x] - \mathrm{E}[W \mid A = a, Z = 0, X = x]$ is not equal to zero for all a, x.

[a]Although an IV must be independent of unmeasured confounders, *A* is a collider between an IV and *U*. Conditioning on the collider will create collider bias such that the IV and *U* become conditionally dependent, thus meeting the conditions for *Z* that are listed in Assumptions 5 and

7.[43,44] Therefore, a valid IV is a special case, meeting the criteria for an NCE even though *Z* is not associated with *U*. Causal effects can be estimated using an IV, an NCO, and double negative control methodology.