U.S. FOOD & DRUG ADMINISTRATION

Center for Biologics Evaluation and Research

Office of Biostatistics and Pharmacovigilance

CBER Surveillance Program Biologics Effectiveness and Safety (BEST) Initiative

BETTER: Bayesian Evaluation of Time-To-Event and Reliability (for vaccine surveillance)

Research Protocol

07/22/2022

Prepared by: Fan Bu and Marc A. Suchard

Author Affiliations: Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA.

Study Team

Authors	Fan Bu, **Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA** Marc A. Suchard, **Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA**
Other Contributors	George Hripcsak, **Department of Biomedical Informatics, Columbia University, New York, NY, USA** Patrick B. Ryan, **Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ, USA** Patricia Lloyd, Hui-Lee Wong, Xinyu Tang, Yun Lu, Richard Forshee, Joyce Obidi, **U.S Food and Drug Administration, Center for Biologics Evaluation and Research, Silver Spring MD, USA**
Version	1.1.3
Delivery Date	07/22/2022

Document History

Version	Date	Description
1.0.1	02/22/2022	Initial draft of protocol
1.1.0	04/01/2022	Restrict study to historical rates methods and Zoster vaccine
1.1.1	05/04/2022	Minor edits based on version 1.1.0
1.1.2	06/28/2022	Minor edits for statistical clarity
1.1.3	07/22/2022	Add clarifying language about statistical methods and quality control

List of Abbreviations

AUC	Area Under the receiver-operator Curve
CCAE	IBM MarketScan Commercial Claims and Encounters
CDM	Common Data Model
CIOMS	Council for International Organizations of Medical Sciences
COVID-19	COronaVIrus Disease 2019
CRAN	Comprehensive R Archive Network
IRB	Institutional review board
LLR	Log Likelihood Ratio
MAP	Maximum A Posteriori
MaxSPRT	MAXimized Sequential Probability Ratio Test
MCMC	Markov Chain Monte Carlo
MDCR	IBM MarketScan Medicare Supplemental Database
MSE	Mean Squared Error
OHDSI	Observational Health Data Science and Informatics

OMOP	Observational Medical Outcomes Partnership
RCT	Randomized controlled trial
WHO	World Health Organization

Contents

1. Protocol Synopsis / Executive Summary

As various approved COVID-19 vaccines are rolled out globally, safety signals have been identified from spontaneous reports and other data sources. The current standard method of safety surveillance adopted by the U.S. Food and Drug Administration (FDA) is MAXimized Sequential Probability Ratio Test (MaxSPRT), which suffers from the inflexibility of a pre-specified sequential analysis schedule. We hope to develop and implement a more flexible Bayesian surveillance framework and compare its performance with MaxSPRT in real-world data.

2. Background and Introductions

Mass vaccination against SARS-CoV-2 is critical to ending the current COVID-19 global pandemic. By the beginning of 2022, 9 vaccines have been approved under the WHO Emergency Use List, and more than 10 billion doses have been administered globally by February 2022 [1]. With the large-scale usage of vaccines under emergency approval, it is essential to ensure their safety and effectiveness through post-market surveillance, as rare but serious adverse events may not be identified in phase 3 clinical trials. In the US, messenger RNA (mRNA) vaccines (BNT162b2, Pfizer-BioNTech; and mRNA-1273, Moderna) were the first SARS-CoV-2 vaccines authorized and as of February 2022, more than 500 million doses of mRNA vaccines have been administered. [2] And yet, there is limited experience with mRNA platforms previous to SARS-CoV-2, and therefore safety surveillance is particularly important to inform public health policy and maintain public trust.

The design of a rapid and reliable vaccine safety surveillance system requires an efficient and robust statistical monitoring approach. The current standard approach used by regulatory agencies in the US is a frequentist sequential analysis method, MaxSPRT [3]. It is designed to control the overall analysis Type I error rate of a sequential analysis by allocating the allowed false positive error over sequential analysis stages. This method has long suffered from its inflexibility as it requires a pre-specified analysis schedule, and, since the alpha spending plan depends on the pre-defined number of data looks, it is prone to prematurely end the analysis and forbids extending the analysis after the pre-chosen analysis endpoint [15].

A more flexible sequential analysis method is, therefore, much desired. A promising candidate is a Bayesian sequential testing framework. Under a Bayesian framework, multiplicity of sequential analyses can be handled more elegantly, without the need for a rigid, pre-specified analysis schedule while allowing continued analyses beyond anticipated endpoints. It is also easier to incorporate historical information into current analyses using a prior distribution through Bayesian inference. With all its theoretical advantages, however, the performance and operating characteristics of Bayesian sequential testing methods have not yet been extensively studied on large-scale real-world data. In particular, with observational health data, Bayesian methods can potentially adjust for unmeasured confounding and sampling bias, but the performance and behavior have not been evaluated in a systematic manner.

The goal of this study is to compare the performance of a Bayesian testing framework with that of MaxSPRT (the current standard approach), in terms of both the hypothesis testing errors (sensitivity and specificity) and estimation accuracy (in estimating the relative risks of adverse events of interest). This study will be conducted using various large-scale health claims databases, in order to understand the operating characteristics in a real-world data-intensive setting. At the initial stage of the study, all analyses will be performed *retrospectively* using historical vaccines with more regular roll-out schedules (such as the Zoster (Shingrix) vaccine). We believe the results of these comprehensive

evaluations will help us better understand the performance and behavior of a Bayesian sequential testing framework and facilitate the design of a more flexible and reliable safety surveillance system for COVID-19 vaccines.

3. Study Objectives

The overarching aim is to compare the performance of frequentist and Bayesian sequential analysis methods for the generation of evidence of vaccine safety in observational, real-world data. Specific aims:

- To evaluate and compare the operating characteristics (Type I and II errors, sensitivity and specificity, etc.) of frequentist and Bayesian sequential testing methods
- To investigate the relationship between threshold choices, Type I and II errors, and time-to-signal of Bayesian sequential methods
- To compare the 'timeliness' of these methods for the identification of vaccine safety signals
- To estimate the bias and precision associated with the use of frequentist and Bayesian methods with the historical rates design for the study of vaccine safety

4. Data Sources

This protocol will be executed on willing data partners across the OHDSI data network. These data sources will encompass a large variety of practice types and populations.

5. Study Design and Definitions

a. Study Design

We will adopt a commonly used design, the historical comparator (historical rates method), using both frequentist (MaxSPRT) and Bayesian sequential testing frameworks. We will conduct the following two types of analyses with variations:

(i) Historical comparator/historical rates (frequentist)
(ii) Historical comparator/historical rates (Bayesian)

To simulate the sequential accrual of data in a sequential analysis setting, we will split the entire exposure period into 12 consecutive analysis periods (each for a one-month interval). This means, for example, at the third analysis time point, **all** data accrued in the first three months since exposure start date will be analyzed.

b. Study Population and Study Period

The study population and study period will vary according to the observational population in health insurance claims records, as described in the data sources table.

c. Exposure Characterization

The evaluation will center on one key vaccine of interest, the Zoster vaccine (Shingrix), with start date 01/01/2018 and end date 12/31/2018, while historical data are retained from 01/01/2017 to 12/31/2017.

Since the Zoster vaccination consists of a two-dose series, we will handle the two doses by further splitting the vaccine exposure to three subset exposures: first dose, second dose, and either dose. The formal cohort definitions of the Zoster vaccine exposure group can be found in Appendix A.

d. Negative Outcome Controls

Negative outcome controls are outcomes believed not to be caused by any of the vaccines, and therefore ideally would not be flagged as a signal by a safety surveillance system. Any effect size estimates for negative control ideally should be close to the null.

A large set of negative outcome controls is defined to representatively represent a wide range of potential systematic errors in an observational study setting. Here, we make the assumption that all negative outcome controls are exchangeable, and, by defining a large set of negative controls, we assume that on average their estimates are close to the true underlying systematic error related to the exposure and data source. To identify negative outcome controls that match the severity and prevalence of suspected vaccine adverse effects, a candidate list of negative controls was generated based on similarity of prevalence and percent of diagnoses that were recorded in an inpatient setting (as a proxy for severity). Manual review of this list by clinical experts created the final list of 93 negative outcome controls [14]. The full list of negative outcome controls can be found in Appendix B.

Negative outcome controls are defined as the first occurrence of the negative control event or any of its descendants in the medical concept vocabulary.

e. Imputed Positive Outcome Controls

Positive outcome controls are outcomes known to be caused by vaccines, and ideally would be detected as signals by a safety surveillance system as early as possible. For various reasons, real positive controls are problematic.[4] Instead, here we will rely on imputed positive outcome controls, created by shifting the estimated effect sizes for the negative outcome controls. We assume the negative outcome controls have a true effect size of 1, so to simulate the estimated effect size when the true effect size is θ we multiply the estimate by θ. For example, if for a negative outcome control a method produces an effect size estimate of 1.1, for a positive outcome control with true effect size of 2 the estimated effect size becomes 1.1 x 2 = 2.2. This approach makes strong assumptions on the nature of the systematic error, most importantly that systematic error does not change as a function of the true effect size. Although this assumption is not likely to hold in the real world, imputing positive outcome controls allows us to provide some indication of what type 2 error to expect for various true effect sizes. For each negative outcome control we will impute positive outcome controls with true effect sizes of 1.5, 2, and 4, so using the 93 negative controls we are able to construct 93 × 3 = 279 positive outcome controls. This increased true effect is applied both for the first and second administrations of the Zoster (Shingrix) vaccine.

f. Outcome of Special Interest – Guillain-Barré Syndrome

In addition to the negative and imputed positive outcome controls, we will further investigate the risk of Guillain-Barré Syndrome (GBS) following the zoster vaccine, as comparison to previous study findings [5]. The previous study by Goud et al. used the self-controlled case series design to

analyze Medicare claims data and found a significant elevated risk (risk ratio 2.34, 95% CI, 1.01-5.41). We will use the historical comparator design, apply both frequentist and Bayesian sequential testing methods, and run analyses on a variety of large-scale databases, in the hope of a more comprehensive analysis of the risk of GBS post zoster vaccination.

g. Metrics

As we will conduct both estimation and testing tasks at the same time, we will compute two sets of metrics based on the study outcomes: (1) metrics for testing, and (2) metrics for estimation. (The following metrics are adapted from previous work [12].)

(i) Testing-related metrics:

- Type 1 error. For negative outcome controls, how often was the null rejected using the various decision rules. This is equivalent to the false positive rate and 1 - specificity.
- Type 2 error. For positive outcome controls, how often was the null **not** rejected using the various decision rules. This is equivalent to the false negative rate and 1 - sensitivity. Will be stratified by true effect size of the positive controls.
- Posterior probability of futility (H_0 true) at final analysis; only reported for Bayesian methods.
- Sensitivity and specificity based on the various decision rules, as well as prior choices in the Bayesian method.
- Detection time: the number of analyses (months) until signals are claimed for 80% of positive controls. This will be stratified by true effect size of the positive controls.
- Rate of contradictory early decisions. For all controls, how often did an earlier signal/futility decision contradict the decision based on full analysis of all data. This can serve as a measure of temporal stationarity of the sequential process — if such contradictory rate is high, then there may be time-varying confounding factors left unadjusted for and further investigation is warranted.
- Rate of "undetermined'' decisions. At each analysis stage, for all controls, how often are the decisions "undetermined" (neither decision thresholds crossed). For most analyses, we would expect this rate to be high at earlier stages when there are not enough data to provide evidence to support either decision but gradually lower as more data accrue.

(ii) Estimation-related metrics:

- Mean precision, computed as 1/(standard error)^2 (for the Bayesian method, "standard error'' is taken as the square root of the posterior distribution variance).
- Mean squared error (MSE). Mean squared error between the log of the effect size point-estimate (MAP estimate for Bayesian method) and the log of the true effect size.
- Area Under the receiver-operator Curve (AUC). This will estimate the ability to discriminate between positive outcome controls and negative outcome controls based on the point estimate of the effect size, which serves as a proxy for both estimation precision and testing accuracy. The AUC will be stratified by true effect size of the control outcomes.
- Coverage. How often the true effect size is within the 95% confidence (or credible) interval.

- Non-estimable. Measure for how many of the negative outcome controls the method was unable to produce an estimate. This typically occurs when the sample size is too small for inference to produce finite estimates.

6. Statistical Methods

Vaccine safety surveillance methods can be broken down into four components: construction of a *counterfactual* (often referred to as the 'expected count'), a *time-at-risk*, the estimation outcome (an *estimator posterior distribution* for the effect size), and a *decision rule* based on the estimation outcome to differentiate signals from non-signals.

a. Counterfactual construction

In this study, we mainly focus on the historical comparator (historical rates) design for counterfactual construction. This design is currently the standard design adopted by various regulatory agencies such as FDA and CDC in the US and the ECDC in the EU.

Traditionally, vaccine surveillance methods compute an expected count based on an incidence rate estimated during some historical time period, for example, in the years prior to the initiation of the surveillance study. [6][7] For Zoster vaccine, we will use the year 2017 (from 2017-01-01 to 2017-12-31) as the historical period and evaluate **four** variations:

- Unadjusted, entire year. Using a single rate computed across the entire historic year for the entire population.
- Age and sex adjusted, entire year. Using a rate stratified by age (in 5-year increments) and sex, computed across the entire historic year (people receiving the Zoster vaccine in 2017 will be excluded from the comparator population). This allows the expected rate to be adjusted for the demographics of the vaccinated.
- Unadjusted, time-at-risk (as defined in the next section) relative to outpatient visit. Using a single rate computed during the time-at-risk relative to a random outpatient visit in the historic year.
- Age and sex adjusted, time-at-risk relative to outpatient visit. Using a rate stratified by age and sex, computed during the time-at-risk relative to a random outpatient visit in the historic year.

For the design variations with age and sex adjustments, age adjustment will be conducted within each data source separately to address age disparities across data sources.

b. Time-at-risk

The time-at-risk is the time window, relative to the vaccination date, when outcomes will potentially be attributed to the vaccine. We define **three** time-at-risk windows: 1-28 days, 1-42 days, and 0-1 days after vaccination.

Time-at-risk windows will be constructed both for the first and second dose for the Zoster vaccine. The time-at-risk for dose one will be censored at the time of the next dose.

c. Estimation Outcome

The effect-size of interest for the historical comparator design is the (log) relative incidence rate ratio. We obtain slightly different estimation outcomes for the frequentist (i.e., MaxSPRT) and Bayesian methods.

For frequentist MaxSPRT, we obtain:

- Effect-size estimate. This is typically a maximum likelihood estimate (MLE) obtained from the analysis.
- Log likelihood ratio (LLR). The log of the ratio between the likelihood of the alternative hypothesis (that there is an effect) and the likelihood of the null hypothesis (of no effect).

The LLR is a convenient and commonly used statistic when performing sequential testing, where the LLR can be compared to a pre-computed critical value, as is done in the MaxSPRT method. [3]

For the Bayesian method, we obtain:

- Posterior distribution for the effect-size, approximated by MCMC posterior samples. This is obtained using Bayes' Rule by combining the likelihood function and the prior distribution. The end result is not a single point estimate but rather a *distribution profile* about our knowledge of the effect-size given accrued data.
- Maximum A Posteriori (MAP) estimate for the effect-size. This is a point estimate obtained by extracting the maxima of the posterior density; this estimate can be regarded as a Bayesian counterpart of the frequentist effect-size estimate.
- Posterior mean for the effect-size. This is a commonly adopted Bayesian estimate, and is, in fact, the optimal Bayesian estimate with squared loss.
- Posterior median for the effect-size. This is another commonly adopted Bayesian estimate and is also the optimal Bayesian estimate with absolute error loss.

For the Bayesian method, we also evaluate different prior distribution choices for the effect-size:

- A log-normal prior with mean =0 and standard deviation (SD) =10 (a **diffuse** prior)
- A log-normal prior with mean =0 and SD =1.5 (a **conservative** prior with >90% mass below 2)
- A log-normal prior with mean =0 and SD =4 (a **weakly informed** prior with ~70% mass below 2)

We choose to use log-normal priors for their simplicity and wide use, in order to focus mainly on comparison between Bayesian and frequentist testing methods. We will consider adopting other prior distributions (e.g., Laplace priors) in subsequent studies.

For both frequentist methods, estimation will be conducted both *with* and *without* empirical calibration. [10,11]. Similarly, for Bayesian methods, analysis will be conducted with and without bias correction based on negative control analysis. Here we assume that the bias induced by

residual systematic error is additive in the log scale of the effect and is invariant to effect size. Therefore, we will estimate a systematic error distribution from the negative outcome control estimates and then use the estimated bias distribution to correct for (potentially biased) estimates.

For the negative outcome controls, empirical calibration will be done in a leave-one-out fashion: when calibrating the estimate for one control, the systematic error distribution will be fitted using all other controls.

d. Decision rule

To identify safety signals, we perform the following statistical hypothesis test regarding the relative rate ratio of an adverse event associated with vaccine exposure:

H_0 : the relative rate ratio of the adverse event ≤ 1 (no elevated risk upon vaccination) v.s.

H_1: the relative ratio ratio of the adverse event > 1 (elevated risk upon vaccination)

To identify 'signals' is to reject the null (H_0) in favor of the alternative hypothesis (H_1). Thus, we need a decision rule, for example, in the terms of a threshold value on one of the estimates statistics.

In our experiment, for the frequentist surveillance method, we will consider a decision rule using the critical value, cv, computed for the log likelihood raio (LLR) at the α=0.05 significance level. That is, we will reject the null and claim a signal when LLR>cv. Here all critical values will be computed using the Sequential package in R, available on R CRAN.

For the Bayesian method, we will implement two sets of decision rules, one for signal (rejecting null) and one for futility/safety (accepting null), by examining the posterior probabilities of the null and alternative hypotheses simultaneously:

- If the posterior probability of signal, P_1=P(H_1 true |data)>δ_1, we claim a signal;
- If the posterior probability of futility/safety, P_0=P(H_0 true |data)>δ_0, we claim safety (non-signal).

We will evaluate three choices of δ_1: 0.80,0.90,0.95, and also three choices of δ_0:0.90,0.95,0.99.

7. Strengths and Limitations

a. Strengths

- Performs a comprehensive evaluation of Bayesian sequential analysis methods for vaccine safety surveillance on large-scale real-world data.
- The fully specified study protocol is being published before analysis begins.
- Dissemination of the results will not depend on estimated effects, avoiding publication bias.
- All analytic methods have previously been verified on real data.
- All software is freely available as open source.

- Use of a common data model allows extension of the experiment to future databases and allows replication of these results on licensable databases that were used in this experiment, while still maintaining patient privacy on patient-level data.
- Use of multiple databases allows estimating consistency to add credibility and supports generalizability.

b. Limitations

- Our follow-up times are limited and variable, potentially reducing power to detect differences in effectiveness and safety.
- We assume hazards are not time varying, and we (at this stage) do not investigate time-varying confounding.
- We only adopt two commonly used study designs (at this stage) which may not be the most suitable design for vaccine safety surveillance situations with complex roll-out schedules (e.g., COVID-19 vaccines).
- Misclassification of study variables is unavoidable in secondary use of health data, so it is possible to misclassify exposure to vaccines, covariates, and outcomes; we do not expect differential misclassification, so bias will most likely be towards the null.

8. Ethical Evaluation

BETTER does not involve human subjects research. As this study involves the retrospective analysis of de-identified and anonymized data, no ethics approval is required. Given that secondary data analysis does not involve personal health information, IRB approval is not needed. Data used in this study were anonymized before its use, and no administrative permissions were required to access raw data. BETTER executes across a federated and distributed data network, where analysis code is sent to participating data partners and only aggregate summary statistics are returned, with no sharing of patient-level data between organizations. All research was performed in accordance with the Declaration of Helsinki guidelines and declarations.

9. Quality Assurance and Control

OHDSI databases use “DataQualityDashboard” as an open-source tool to evaluate data quality. All analyses are performed in R, including using OHDSI HADES packages, which have been extensively used in prior publications and are validated through unit test cases.

10. References

[1] WHO. WHO coronavirus (COVID-19) dashboard. 2022.https://covid19.who.int/ (accessed 14 Feb 2022).

[2] Statista. COVID-19 vaccinations administered in the US, by vaccine manufacturer. 2022.https://www.statista.com/statistics/1198516/covid-19-vaccinations-administered-us-by-company/ (accessed 14 Feb 2022).

[3] Kulldorff M, Davis RL, Kolczak† M, *et al.* A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Analysis* 2011;**30**:58–78. doi:10.1080/07474946.2011.539924

[4] Schuemie MJ, Ryan PB, Hripcsak G, *et al.* Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philosophical transactions Series A, Mathematical, physical, and engineering sciences* 2018;**376**.

[5] Goud R, Lufkin B, Duffy J, *et al.* Risk of guillain-barré syndrome following recombinant zoster vaccine in medicare beneficiaries. *JAMA internal medicine* 2021;**181**:1623–30.

[6] Li X, Ostropolets A, Makadia R, *et al.* Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: Multinational network cohort study. *bmj*2021;**373**.

[7] Klein NP, Lewis N, Goddard K, *et al.* Surveillance for adverse events after COVID-19 mRNA vaccination. *JAMA* 2021;**326**:1390–9.

[8] Whitaker HJ, Farrington CP, Spiessens B, *et al.* Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;**25**:1768–97.

[9] Glanz JM, McClure DL, Xu S, *et al.* Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *J Clin Epidemiol* 2006;**59**:808–18.

[10] Schuemie MJ, Ryan PB, DuMouchel W, *et al.* Interpreting observational studies: Why empirical calibration is needed to correct p-values. *Statistics in medicine* 2014;**33**:209–18.

[11] Schuemie MJ, Hripcsak G, Ryan PB, *et al.* Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences of the United States of America* 2018;**115**:2571–7.

[12] Schuemie MJ, Cepeda MS, Suchard MA, *et al.* How confident are we about observational findings in health care: A benchmark study. *Harvard Data Science Review* 2020;**2**.

[13] Woelfle M, Olliaro P, Todd MH. Open science is a research accelerator. *Nature chemistry* 2011;**3**:745–8.

[14] Shuemie MJ, Arshad F, Pratt N, *et al.* Vaccine safety surveillance using routinely collected healthcare data – An empirical evaluation of epidemiological designs. *medRxiv* 2021.08.09.21261780.

[15] Li R, Stewart B, Rose C. A Bayesian approach to sequential analysis in post-licensure vaccine safety surveillance. Pharmaceutical Statistics. 2020 May;19(3):291-302

11. Appendices

A. Exposure Cohort Definitions: Zoster Vaccines

A.1 Cohort Entry Events

People enter the cohort when observing any of the following:

1. drug exposures of 'Shingrix,' starting between January 1, 2018 and December 31, 2018.

A.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 0 days.

A.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

A.4 Concept set: Shingrix

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
792784	varicella zoster virus glycoprotein E Injection [Shingrix]	1986828	RxNorm	NO	YES	NO|
792783	varicella zoster virus glycoprotein E, recombinant 0.1 MG/ML [Shingrix]	1986827	RxNorm	NO	YES	NO|

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
792788	varicella zoster virus glycoprotein E, recombinant 0.1 MG/ML Injection [Shingrix]	1986832	RxNorm	NO	YES	NO|
36421491	Varicella-Zoster Virus Vaccine Live (Oka-Merck) strain Injectable Solution [Shingrix]	OMOP4763774	RxNorm Extension	NO	YES	NO|
792785	Shingrix Injectable Product	1986829	RxNorm	NO	YES	NO|
706103	zoster vaccine recombinant	187	CVX	NO	YES	NO|

B. Negative controls

As described in Section 5, 93 negative outcome controls were manually selected by domain experts (see [14] for details). Some of these negative outcomes only occur in one sex group, which can be treated just as regular negative controls in a sex-adjusted design. In an unadjusted design, however, this would indicate extreme differential gender misclassification. Given that the goal of having a large set of negative outcome controls is to cover a wide population of possible systematic errors (including differential gender misclassification), we have chosen to retain these gender-specific negative controls, as they could help us gauge gender-specific biases that might otherwise get neglected. At the same time, we have taken such potential misclassification into account in applying them. Also, since fewer than 10% of all negative control outcomes are gender-specific, they would not excessively affect the overall results.

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
438945	Accidental poisoning by benzodiazepine-based tranquilizer
434455	Acquired claw toes
316211	Acquired spondylolisthesis
201612	Alcoholic liver damage
438730	Alkalosis
441258	Anemia in neoplastic disease
432513	Animal bite wound
4171556	Ankle ulcer
4098292	Antiphospholipid syndrome
77650	Aseptic necrosis of bone
4239873	Benign neoplasm of ciliary body
23731	Benign neoplasm of larynx
199764	Benign neoplasm of ovary

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
195500	Benign neoplasm of uterus
4145627	Biliary calculus
4108471	Burn of digit of hand
75121	Burn of lower leg
4284982	Calculus of bile duct without obstruction
434327	Cannabis abuse
78497	Cellulitis and abscess of toe
4001454	Cervical spine ankylosis
4068241	Chronic instability of knee
195596	Chronic pancreatitis
4206338	Chronic salpingitis
4058397	Claustrophobia
74816	Contusion of toe

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
73302	Curvature of spine
4151134	Cyst of pancreas
77638	Displacement of intervertebral disc without myelopathy
195864	Diverticulum of bladder
201346	Edema of penis
200461	Endometriosis of uterus
377877	Esotropia
193530	Follicular cyst of ovary
4094822	Foreign body in respiratory tract
443421	Gallbladder and bile duct calculi
4299408	Gouty tophus
135215	Hashimoto thyroiditis
442190	Hemorrhage of colon

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
43020475	High risk heterosexual behavior
194149	Hirschsprung's disease
443204	Human ehrlichiosis
4226238	Hyperosmolar coma due to diabetes mellitus
4032787	Hyperosmolarity
197032	Hyperplasia of prostate
140362	Hypoparathyroidism
435371	Hypothermia
138690	Infestation by Pediculus
4152376	Intentional self poisoning
192953	Intestinal adhesions with obstruction
196347	Intestinal parasitism
137977	Jaundice

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
317510	Leukemia
765053	Lump in right breast
378165	Nystagmus
434085	Obstruction of duodenum
4147016	Open wound of buttock
4129404	Open wound of upper arm
438120	Opioid dependence
75924	Osteodystrophy
432594	Osteomalacia
30365	Panhypopituitarism
4108371	Peripheral gangrene
440367	Plasmacytosis
439233	Poisoning by antidiabetic agent

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
442149	Poisoning by bee sting
4314086	Poisoning due to sting of ant
4147660	Postural kyphosis
434319	Premature ejaculation
199754	Primary malignant neoplasm of pancreas
4311499	Primary malignant neoplasm of respiratory tract
436635	Primary malignant neoplasm of sigmoid colon
196044	Primary malignant neoplasm of stomach
433716	Primary malignant neoplasm of testis
133424	Primary malignant neoplasm of thyroid gland
194997	Prostatitis
80286	Prosthetic joint loosening
443274	Psychostimulant dependence

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
314962	Raynaud’s disease
37018294	Residual osteitis
4288241	Salmonella enterica subspecies arizonae infection
45757269	Sclerosing mesenteritis
74722	Secondary localized osteoarthrosis of pelvic region
200348	Secondary malignant neoplasm of large intestine
43020446	Sedative withdrawal
74194	Sprain of spinal ligament
4194207	Tailor’s bunion
193521	Tropical sprue
40482801	Type II diabetes mellitus uncontrolled
74719	Ulcer of foot
196625	Viral hepatitis A without hepatic coma

Table B.1: Negative outcome controls.

Outcome Id	Outcome Name
197494	Viral hepatitis C
4284533	Vitamin D-dependent rickets

C. Additional investigated outcome cohort

C.1 Adverse Event Outcome - Guillain Barré Syndrome

C.1.1 Cohort Entry Events

People may enter the cohort when observing any of the following:

1. condition occurrences of ‘Guillain-Barré syndrome.’

Restrict entry events to having at least 1 visit occurrence of ‘Inpatient or Inpatient/ER visit,’ starting anytime on or before cohort entry start date and ending between 0 days before and all days after cohort entry start date.

C.1.2 Inclusion Criteria

C.1.2.1 1. has no events in prior ‘clean window’

Entry events having no condition occurrences of ‘Guillain-Barré syndrome’, starting in the 365 days prior to cohort entry start date, having at least 1 visit occurrence of ‘Inpatient or Inpatient/ER visit’, starting anytime on or before ‘Guillain-Barré syndrome’ start date and ending between 0 days before and all days after ‘Guillain-Barré syndrome’ start date. Any prior history of ‘Guillain-Barré syndrome’, even more than 365 days before the relevant cohort entry start date, is also considered prior occurrence and thus will result in exclusion.

C.1.3 Cohort Exit

The cohort end date will be offset from index event's start date plus 1 day.

C.1.4 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

C.1.5 Concept set: Guillain-Barré syndrome

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
374925	Acute infective polyneuritis	129131007	SNOMED	NO	YES	NO
4164770	Guillain-Barre syndrome	40956001	SNOMED	NO	YES	NO
4070552	Fisher's syndrome	1767005	SNOMED	NO	YES	NO

C.1.6 Concept set: Inpatient or Inpatient/ER visit

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
262	Emergency Room and Inpatient Visit	ERIP	Visit	NO	YES	NO
9201	Inpatient Visit	IP	Visit	NO	YES	NO