



Utrecht
University

COVID-19 VACCINES AND THE MISCLASSIFICATION OF ADVERSE EVENTS

Riikka Holtari

Thesis submitted to

Utrecht University

for the degree of

MASTER OF SCIENCE

Faculty of Medicine

Graduate School of Life Sciences

Epidemiology,

2023

With thanks to supervisors:

Professor Dr Rene Eijkemans, PhD

Dr Ivonne Martin, PhD

ABSTRACT

In order to evaluate the impact of outcome misclassification on the possible causal association between Adverse Events of Special Interest (AESIs) and COVID-19 vaccination, we conducted a literature review and a simulation study. The literature study aimed to obtain a plausible range of outcome misclassification indices of the International Classification of Diseases (ICD) coding systems used in electronic healthcare databases.

We used logistic regression to contrast a naïve estimator that disregards misclassification with a misclassification-integrating Maximum Likelihood Estimation (MLE) model to explore the relationship between vaccine exposure and the occurrence of AESIs. The MLE model employed marginal probability from a Bernoulli distribution to account for misclassification, facilitating a comparison of log of odds ratios and relative risks between the models.

In our simulation study, we generated data which incorporated a fixed vaccination rate, varying sample sizes, regression coefficients, and misclassification rates to evaluate bias and mean squared error (MSE) of the log of odds ratio and the relative risk. The analyses showed that the MLE model exhibited reduced bias under high prevalence and specificity conditions when examining the relative risk bias. Despite its great variability, the MLE model outperformed the naïve estimator in specific simulations with increased association strengths, supporting our hypothesis. Our findings reveal a need for rigorous methodologies to address misclassification in vaccine safety assessments and indicate that the degree of misclassification may be significant, depending on the specific ICD code.

Our research highlighted the importance of thoroughly recognising misclassified data and outlined critical areas for future research in epidemiology. Our observations demonstrated that the traditional approach, which disregards the subsequent bias brought on by misclassification – is not fundamentally flawed, emphasising the need to investigate how and when these errors can be significant.

Words: 285

Keywords: Misclassification, Adverse Events of Special Interest (AESI), COVID-19 vaccination, Simulation study, International Classification of Diseases (ICD), Bias, Mean Squared Error (MSE).

LAYMAN'S SUMMARY

This research paper explored the misclassification of Adverse Events of Special Interest (AESIs) after COVID-19 vaccination. These are pre-selected adverse events extracted from the *COVID-Vaccine-Monitor: Rapid Safety Assessment of SARS-CoV-2 vaccines in the EU Member States using electronic health care data sources* -protocol. In order to evaluate the impact of outcome misclassification on the relationship between COVID-19 vaccination and the AESIs, we conducted a comprehensive literature review followed by an in-depth simulation study.

Our literature review revealed misclassification errors in electronic healthcare databases which use the International Classification of Diseases (ICD) coding systems. Across the review of misclassification rates, we detected that some AESIs are more prone to misclassification, and the degree of misclassification can be measured using metrics such as sensitivity and specificity.

To examine the impact of misclassification on the interpretation of vaccine safety, we analysed two logistic regression models. The naïve model simulated the traditional approach and did not factor in misclassified outcomes, while the Maximum Likelihood Estimation (MLE) model did. Unlike traditional approaches, the MLE model accounts for misdiagnosed AESI codes within an electronic database. We compared our naïve estimator to our MLE model, uncovering factors that bias the association between vaccine exposure and AESIs. While both outcome and exposures are assumed to be binary variables, the MLE model used the Bernoulli distribution.

Our simulation study involved synthesising data with a fixed vaccination rate, varying sample sizes, effect sizes, and misclassification rates. These analyses revealed that the MLE estimator displayed a reduced bias for relative risk when the prevalence of a given AESI was high and the reported misclassification rate of the ICD code was low. However, when the prevalence was low, and the strength of association was weak, the MLE did not show superior accuracy in all simulations. Nevertheless, the MLE model's greater variability but reduced bias showed that it outperformed the naïve estimator in specific simulations, especially those with a stronger simulated association – supporting our hypothesis.

Our findings expose a need for a deeper understanding of the underlying influences affecting the bias and error introduced by the inadvertent misclassification of AESIs. Moreover, the study stresses the need for rigorous methodologies to address misclassification in future vaccine safety studies. Arguably, our observations relay that the traditional approach is not fundamentally flawed; given its disregard for misclassified data, this oversight does not always result in bias and inaccuracy in healthcare databases.

Words: 395

TABLE OF CONTENTS

Introduction.....	6
Literature Review.....	7
Methodology for Literature Review	7
Inclusion Criteria	7
Exclusion Criteria	7
Quality Assessment Questions.....	7
Results of the Literature Review	8
Autoimmune Diseases.....	8
Cardiovascular System.....	8
Circulatory System.....	9
Hepato-gastrointestinal and Renal System	10
Nerves and Central Nervous System Disorders	10
Respiratory System	11
Skin and Mucous Membrane, Bone and Joints System	11
Other Systems	11
Maternal Pregnancy Outcomes	12
Neonatal Pregnancy Outcomes	13
Methods	14
Logistic Regression	14
Methodology for Simulation study.....	16
Results.....	18
Bias Analysis for the Estimated Log of Odds	18
Bias Analysis for the Estimated Relative Risk	21
Mean Squared Error Analysis for the Estimated Log of Odds Ratio	21
Mean Squared Error Analysis for the Estimated Relative Risk.....	23
Discussion	24
Conclusion	27
Supplementary Materials	32
Appendix 1: Literature Review Table Summary	32
Appendix 2: RStudio Script for the Calculation of Minimum and Maximum Sensitivity..	35
Appendix 3: RStudio Packages Used in Simulation Study.....	37
Appendix 4: The Complete RStudio-Script for the Simulation Study.....	38
Appendix 5: Bias Analysis for the Estimated Relative Risk	43
Appendix 6: Mean Squared Error Analysis for the Estimated Relative Risk.....	45

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 outbreak in 2019 and the resulting coronavirus disease (COVID-19) catalysed a mass vaccination campaign globally (Graham, 2020; Bok et al., 2021). Implementing an immunisation program of this magnitude demands a greater degree of scrutiny regarding vaccine safety. While vaccine development typically spans decades, the development of the COVID-19 vaccine was expedited to a matter of months, placing even greater importance on the surveillance of vaccine adverse events. In order to assess the occurrence of adverse events after vaccine exposure, we must rely on accurate approximations of their background rates. Therefore, it is crucial to consider the validity of reported adverse events and their diagnoses – significant misclassification could potentially weaken the assessment of vaccine associations (Graham, 2020; Patel et al., 2020).

This research paper is part of the project: COVID-Vaccine-Monitor: *Rapid Safety Assessment of SARS-CoV-2 vaccines in the EU Member States using electronic health care data sources*. It focuses on the pre-selected Adverse Events of Special Interest (AESIs) associated with vaccine exposure. The study's data retrieval and transformation were based on case-recognising algorithms using the 9th and 10th revisions of the World Health Organisation's International Classification of Diseases (ICD) system. The ICD system is a diagnostic tool used for comparing morbidity and mortality statistics, insurance, and reimbursement purposes. Linking each AESI to ICD-9 and ICD-10 codes allows for the evaluation of misclassification of AESI diagnoses to conduct a valid assessment of vaccine safety (Mullooly, 2008).

The ICD-9 and ICD-10 systems are currently the most common versions used by world healthcare databases. The validity of ICD codes is assessed by evaluating misclassification measures such as the Positive Predictive Value (PPV) – the proportion of true positive cases relative to the total positives identified by the electronic system. By comparing the reported PPVs of different codes through a literature review, we can assess the relative validity of the same and determine the reliability of diagnoses. Therefore, this paper aims to provide critical insights into the accuracy of adverse event reporting post-COVID-19 immunisation; thus, we hypothesise that proper accounting for misclassification will be meaningful for the safety evaluation of novel vaccine technology.

In response, our research question is: what is the association between misclassification errors and the reporting of AESIs following COVID-19 vaccination? The following findings could be critical in informing public health policies concerning clinical decision-making related to COVID-19 vaccinations. Furthermore, this research addresses a gap within the pre-existing literature and highlights the need for a more in-depth understanding of outcome classification within electronic healthcare databases. By exploring this question, this paper will significantly contribute to the ongoing discussion around the safety of COVID-19 vaccines.

Literature Review

Validating ICD algorithms within electronic healthcare databases for real-world evidence studies is increasingly important. Poorly validated ICD algorithms can lead to inaccurate or misleading diagnoses. A proposed framework for validating ICD algorithms includes several steps, such as defining the study population, selecting the appropriate ICD codes, and assessing the algorithm's performance using various metrics (Tanpowpong, 2020).

The literature review compiles the available misclassification rates by finding the PPVs for each AESI and its code. The ICD codes [found in **Appendix 1**] were extracted from the *Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines-report* and refer to their respective preliminary reports found on the Zenodo platform, an open-access repository for sharing research data and outputs.

Methodology for Literature Review

We used the Boolean operator 'AND' to combine three search concepts for each disease and condition: the specific disease or condition, the respective ICD code, and the concepts of validity and positive predictive value (PPV). For example, for Guillain Barre Syndrome, we combined 'Guillain Barre Syndrome', 'ICD-9', '357.0', and the validity keywords (including 'PPV', 'positive predictive value', 'validity', 'validation') using 'AND'. Similarly, for Acute Disseminated Encephalomyelitis, we combined 'Acute Disseminated Encephalomyelitis', 'ICD-10', 'G04.0 OR G04.8', and the validity keywords. This strategy was consistent for all the conditions and their respective ICD codes: 'Acute Aseptic Arthritis' and '274'; 'Diabetes Mellitus' and '250.x'; 'Idiopathic Thrombocytopenic Purpura (ITP)' and 'D69.3 OR 287.31'; 'Microangiopathy' and '448.0 OR I78'; 'Heart Failure' and 'I50.20 TO I50.23'; 'Coronary Artery Disease' and 'I20.0 OR I21 OR I24'; 'Ventricular Arrhythmia' and '427.1 OR 427.4 OR 427.5'; 'Myocarditis' and '422 OR I40 OR I41'; and so on for all the remaining diseases and conditions including COVID-19, various pregnancy-related conditions and infant conditions [found in **Appendix 1**].

Inclusion Criteria

1. The study was published between 1 January 1995 and 1 March 2023.
2. The study has been written in English.
3. The study used medical data validating ICD code assignment.
4. The study fulfilled our Quality Assessment Questions.

Exclusion Criteria

1. The study focuses on diagnostic test accuracy.

Quality Assessment Questions

1. Was the research question clearly stated?
2. Was the study design appropriate for answering the research question?
3. Was the sample size adequate and representative?
4. Were the data collection methods valid and reliable?
5. Were the data analysis methods appropriate and transparent?

6. Did the study define a study population, select appropriate ICD codes, and use appropriate metrics to assess algorithm performance?

The values reported are primarily PPVs from various scientific papers. This report splits them into their respective body systems, focusing on the general outcomes for the general population.

Results of the Literature Review

Autoimmune Diseases

Guillain Barre Syndrome

Guillain Barre Syndrome (GBS) is typically identified using ICD-9 code 357.0. PPVs in studies have varied, ranging from 12%-16% in a PRISM study to 29% in a study by Funch et al. (2013). However, the PPV increases to 82% when the definition is restricted to primary inpatient diagnosis codes (Burwen et al., 2012).

Acute Disseminated Encephalomyelitis

Acute Disseminated Encephalomyelitis (ADEM) is designated by the ICD-10 code G04.0. A Danish NRP study, which also included G04.8 and G04.9, reported a PPV of 44% for ADEM (Boesen et al., 2018).

Acute Aseptic Arthritis

Gout, encoded as 274 in ICD-9, showed a PPV of 61% in a study by Harrold et al. (2007). The study also reported relatively low PPVs for Osteoarthritis and Rheumatoid Arthritis (60% and 59%, respectively).

Diabetes Mellitus

For Diabetes Mellitus, the ICD-9 codes 250.x1 for Type I and 250.x0 for Type II have been studied with high PPVs ranging from 94% to 97% based on two or more indicators (Zgibor et al., 2007). Khohkar's (2016) study presented a wider range of sensitivity, specificity, and PPVs using combined physician claims data and hospital discharge records.

Thrombocytopenia

Idiopathic Thrombocytopenic Purpura (ITPa) is identified using ICD-10 code D69.3. Studies reported high PPVs, with values of 93% and 95.8%. However, the ICD-9 code 287.3 for primary thrombocytopenia showed lower PPVs, ranging from 65% to 85%. After updating the code to 287.31 in 2006 for immune thrombocytopenic purpura (ITPb), the PPV in inpatient settings rose to 80% (Galdarossa et al., 2012; Heden et al., 2009).

Cardiovascular System

Microangiopathy

Microangiopathy validation studies are scarce. Gillmeyer et al. (2019) included ICD codes 448.0 and ICD-10 I78 for pulmonary capillary hemangiomas (PCH), but no specific PPV was reported. The best-performing algorithm for identifying pulmonary arterial hypertension (PAH) required ICD diagnosis codes, right heart catheterisation (RHC) codes, and PAH-specific therapy (VA hospitals: PPV, 70.0%; BMC: PPV, 86.0%).

Heart Failure

Pocobelli et al. (2022) validated the ICD-10 codes for heart failure (HF) and its subtypes. Diagnostic codes I50.20-I50.23 have a PPV of 41.4%. The highest PPV was for HF patients with mid-range ejection fraction (mEF) or preserved ejection fraction (pEF), with diastolic HF at a PPV of 92.0% and pEF/mEF at 97.7%.

Stress Cardiomyopathy

Bhat et al. (2020) reported a PPV of 98% for stress cardiomyopathy (SCM), using both ICD-9 (429.83) and ICD-10 (I51.81) codes. The PPV rose to 100% when only considering patients with a primary diagnosis of SCM.

Coronary Artery Disease

Bezin et al. (2015) studied acute coronary syndrome and found the highest PPVs for individual codes I20.0, I21, and I24. The combination of I20.0, I21 or I24 had the best performance for identifying acute coronary syndrome, with a PPV of 84.2%.

Arrhythmia

Ye et al. (2018) validated algorithms to identify ventricular arrhythmia and sudden cardiac death. They reported an algorithm identifying ventricular arrhythmia and used ICD-9 codes 427.1, 427.4, and 427.5. This study validated the use of the codes with a PPV of 93%. A variation of this algorithm, including code 427.69, lowered the PPV to 82%. Adding codes for unspecified cardiac arrhythmia further lowered the PPV to 50%.

Myocarditis and Pericarditis

Sundbøll et al. (2016) reported a PPV of 64% for myocarditis and 92% for pericarditis in the Danish National Patient Register. The PPVs were higher in university hospitals and younger patients. The authors concluded that these diagnoses require caution when used for research due to their relatively lower PPVs.

Circulatory System

Disseminated intravascular coagulation

In a study by Wada et al. (2014), ICD-10 codes for disseminated intravascular coagulation (DIC) were validated, and they reported PPVs ranging from 68.2% to 90.3%. They used medical records as the standard, indicating that ICD codes for DIC can help identify cases with a relatively low risk of misclassification.

Venous thromboembolism

Öhman et al (2018) validated ICD8/9/10 codes for pulmonary embolisms (PE) and Deep Vein Thrombosis (DVT), with an overall PPV of 71.1%. The PPV was 85.8% for PE and 54.1% for DVT. The authors found that approximately 9% of PE and DVT events are misclassified, predominantly as DVT events.

Stroke

In a systematic review by McCormick et al (2015), acute stroke ICD codes were found to accurately predict true cases of any type of acute stroke and of particular subtypes. For ischaemic stroke, identified by ICD-9 434 or ICD-10 I63, the PPV was 82%.

Cerebral venous thrombosis

Handley and Emsley (2020) reported a validation of ICD-10 codes for intracranial venous thrombosis and showed an overall PPV of 92.3% for the codes G08.X, O22.5, I67.6, I63.6, O87.3.

Hepato-gastrointestinal and Renal System

Acute Liver Injury

The European Association characterises Acute Liver Failure (ALF) as a rare and highly specific syndrome, typically indicated by abnormal liver blood tests in the absence of underlying chronic liver disease (CLD) (McDowell Torres et al., 2010). In a study by Lo Re et al. (2013), the positive predictive values (PPVs) of individual ICD-9 codes for severe acute liver injury (SALI) varied significantly, from 6.5% to 54.3%. However, when specific codes were observed in conjunction, the PPV improved, with the combination of 573.3 and 570 reaching 60%. Due to the specificity of the codes and the rarity of ALF, the PPVs for true ALF were relatively low, between 1.8% and 9.1%. However, in patients without CLD, the PPV for codes 570 and 572.8 was 100%. Another study investigating coding algorithms following acetaminophen overdose found a PPV of 85% for hepatotoxicity and 53% for ALF in patients with acetaminophen hepatotoxicity (Myers et al., 2007).

Acute Kidney Failure

Waikar et al. (2006) reported that the most common ICD-9 codes for Acute Kidney Failure (AKF) were 584.5 and 584.9, and the codes had a PPV of 80.2%.

Nerves and Central Nervous System Disorders

Generalised Convulsions and Epilepsy

In a study by Jetté et al. (2010), ICD-9 epilepsy codes were validated. Code 345.x was examined alongside diagnoses resembling epilepsy, such as codes 346.x, 435.x, 780.28, and 780.3. The ICD-10 epilepsy code is G40.x or G41.x, with similar resembling diagnoses included. The PPV for seizure monitoring unit (SMU) database records reported 84.9% overall, with varying PPVs for specific codes. The ICD-9 code 345's validity study showed high PPVs across care settings, with lower PPVs when combined with convulsion diagnostic code 780.3. ICD-10 PPVs for G40-41 followed a similar pattern.

The given sequence shows PPVs rising in relation to patient cohort nature. In Shui et al.'s paper (2009), ICD-9 codes' PPVs were assessed for identifying seizures in children six weeks to 23 months post-vaccination. Codes 730.8 and 345 had the highest PPVs in emergency room settings, lower in inpatient settings, and lowest in outpatient settings. They concluded that PPVs might suffer less bias when restricted to ER diagnoses.

In Gmuca et al. (2020), code 341.0 had a 47% PPV across all sites. However, the most successful case-finding algorithms included at least five codes, resulting in a 90% PPV in children and 92% in adults when measured through patients with documented hospitalisation.

Transverse Myelitis

Boesen et al (2018) reported PPVs for children with G37.3 [transverse myelitis] as 64%. They concluded that the PPVs were acceptable for transverse myelitis and that using the ICD-10 code is beneficial for epidemiological studies.

Respiratory System

Acute Respiratory Distress Syndrome

In a study by Thomsen & Morris (1995), the positive predictive values (PPVs) for the ICD-9 discharge diagnosis codes 518.5, 518.81, and 518.82, pertaining to acute respiratory distress syndrome (ARDS), were found to be low. In all six of the surveyed hospitals, PPVs were recorded to be 7%. The authors attributed these low PPVs to the low incidence of ARDS in the hospitals, coupled with the challenges associated with ICD-9 coding for ARDS. They noted that a low PPV is expected when screening for rare conditions like ARDS.

COVID-19

Regarding COVID-19, the ICD-10 code is U07.1, several studies have evaluated the validity of this code. As reported by Lynch et al. (2021), out of 52,000 hospitalisations in April 2020, the PPV for code U07.1 was found to be 91.52%. Kluberg et al. (2022) reported a PPV of 93% for the code, while Bhatt et al. (2021) found a PPV of 90.0% among 1,208 patients. These results suggest that code U07.1 has high PPVs, making it valid for identifying COVID-19 cases in inpatient data. As the pandemic progressed, the use of U07.1 and R05 [Cough] codes declined, and diagnosis codes Z20.828 and Z11.59 became more frequently used.

Skin and Mucous Membrane, Bone and Joints System

Erythema Multiforme

Erythema Multiforme (EM), also known as Steven-Johnsons Syndrome (SJS), is represented by ICD-9 code 695.1x. A study by Eisenberg et al. (2012) found the positive predictive values (PPVs) for this condition to be notably low, with only 2.00% recorded for inpatient claims. Another study by Schneider et al. (2012) investigated the incidence of EM and SJS, employing ICD-8 and ICD-9 codes for validation. The ICD-8 code 695.1 showed a PPV of 59.6% based on a review of discharge summaries, hospital records, and outpatient data. In contrast, ICD-9 code 695.1, which refers to both SJS and EM, revealed a PPV of 54.0% in one validation study and 53.7% in another, both based on medical record reviews.

Chilblain like lesions

There is currently no available literature concerning the validation of ICD-9 or ICD-10 codes for chilblain-like lesions, a skin condition associated with COVID-19. A study by Kanitakis et al. (2020) on this subject only included 17 patients, and no ICD validation results were found.

Levine et al. (2013) demonstrated that ICD-9 codes can be used to retrospectively identify skin and soft tissue infections (SSTIs) with high certainty. However, this accuracy diminishes when more data like microbiology information and CPT codes are incorporated. The study found that the identification of cellulitis/abscess using ICD-9 codes had a PPV of 91.5% (95% CI: 88.9-94.1%). In contrast, the identification of cellulitis/abscess specific to a particular body site had a PPV of 52.6% (95% CI: 47.6-57.6%). The PPVs were generally lower when the SSTI identification algorithm was restricted to initial visits. More research is needed to evaluate the sensitivity of this method.

Other Systems

Anosmia, dysgeusia

Anosmia and dysgeusia have minimal validation for their ICD codes, but a study by Zayet et al. (2021) found that these symptoms had a positive predictive value (PPV) of 77% for a positive RT-PCR result for COVID-19. The combination of these two symptoms had a PPV of 83% for a

positive SARS-CoV-2 RT-PCR result, making them valuable indicators of COVID-19 infection in outpatients.

Anaphylaxis

For anaphylaxis, a study by Mesfin et al. (2019) validated the use of ICD-10 codes T80.5, T80.6, T88.1, T88.6, and T78.2 in identifying cases of anaphylaxis after vaccination. The study revealed a low overall PPV, with only 24.6% of the 69 records confirmed as anaphylaxis due to vaccination. However, when using the ICD-10 code T80.5, the PPV was 95.7% for all anaphylaxis cases and 52.2% for anaphylaxis due to vaccination.

A study by de Sordi et al. (2021) also evaluated anaphylaxis identification using ICD-10 codes. They identified potential cases by combining specific in-hospital codes, outpatient codes with symptom codes, or non-specific in-hospital codes with two symptom codes. The main algorithm had a PPV of 62.8% based on 49 true cases, but it underestimated absolute risks by about a third. Restricting the algorithm to primary discharge codes improved PPV but reduced sensitivity.

Multisystem Inflammatory Syndrome in Children (MIS-C)

Multisystem Inflammatory Syndrome in Children (MIS-C) is a rare condition linked to COVID-19 infection. An emergency ICD-10 code was introduced in 2020 to account for MIS-C, but it is unclear if this code accurately identifies patients with COVID-19. An early study found that only 20.3% of known COVID-19 cases were assigned this ICD-10 code, though the PPV of the U07.1 code was 96.3% (Aldawas & Ishfaq, 2022).

Death

Death is not associated with an ICD code, but sudden death is linked to the ICD-9 code 798 and ICD-10 codes R99, R95, R96.0, and R96.1. However, these codes for sudden death have not been validated, as per a study by Khalili et al. (2012). The code R99, for ill-defined causes of death, is typically attributed to a specific cause.

Maternal Pregnancy Outcomes

Gestational Diabetes

In the case of gestational diabetes, the ICD-9 code 648.8 had a reported positive predictive value (PPV) of 11.6%, as per Donovan et al.'s 2019 study.

Preeclampsia

For preeclampsia, Watson (2016) et al. reported a PPV of 81.2% for the ICD-10 O14 code. Geller et al. reported a lower PPV (45.3%) for mild or unspecified preeclampsia (ICD-9 642.4) but a higher PPV (84.8%) for severe preeclampsia.

Maternal death

Regarding maternal death, Sigakis et al. (2016) reported a range of PPVs for various ICD-9 codes capturing severe maternal morbidity during delivery hospitalisation. The PPV was high when supported by objective evidence, such as laboratory values or procedure documentation. However, the PPV was low for codes that required appropriate judgment and interpretation of clinical data, like the category 'Severe anaesthesia complications'. As such, these codes should be used cautiously in administrative research compared to codes primarily defined by objective data. However, specific PPVs for maternal death were not reported.

Neonatal Pregnancy Outcomes

Foetal growth restriction

Phiri's (2016) study on foetal growth restriction (FGR) reported that the ICD-9 code 656.5x for small for gestational age (SGA) infants had a low sensitivity (14.2%) but high specificity (99.7%) and positive predictive value (PPV, 86.8%). A systematic review by Watson et al. (2021) found a PPV of 57.1% for FGR using the GAIA case definition, suggesting caution in interpreting results.

Spontaneous Abortions

In the context of spontaneous abortions, Chomistek et al. (2023) found high PPVs for several claims-based algorithms, with the most uncomplicated having a PPV of 84.7% (95% CI 78.3, 91.2) using ICD-10 code O03.

Stillbirth

Andrade et al. (2021) showed that an ICD-10-based algorithm could effectively identify stillbirth events, with the highest PPV of 82.5% achieved when a diagnosis code indicated a gestational age ≥ 20 weeks.

Preterm-birth

Using the GAIA case definition, Watson et al (2021) found a PPV of 75.5% for preterm birth (PTB). They also reported PPVs of 56.8 (47.7–65.8) for the code O60 and 76.4 (67.6–85.2) for P07.

Major Congenital Anomalies

Chomistek et al. (2023) reported that an algorithm had a PPV of 44.0% for identifying significant congenital anomalies, while another, which required two claims separated by at least 30 days, had a PPV of 67.8%.

Microcephaly

For microcephaly, Watson et al. (2021) reported a PPV of 40.0 (29.0–50.0) for the ICD-10 code Q02.

Neonatal Death

A study by Gray et al. (2022) on neonatal death reported a sensitivity of 65.5% and a PPV of 90% using the ICD-10 code for cardiac arrest.

Termination of Pregnancy for Foetal Anomaly (TOPFA)

In Garne et al.'s study (2023) on termination of pregnancy for foetal anomaly (TOPFA), 99% of cases were identified in hospital databases, with 91% having an end-of-pregnancy code and 82% having a code for a congenital anomaly. Salemi et al. (2018) reported a PPV of 94.2% for TOPFA using the Florida Birth Defects Registry, which increased with more restrictive case definition algorithms.

Methods

The primary aim of most vaccine safety studies is to establish a causal relationship between vaccination and AESIs. Typically, the gold standard design used to infer causality is a randomised controlled trial; however, the potential presence of misclassification errors can affect the determination of causality.

Logistic Regression

Researchers have frequently ignored misclassification, even though this might have introduced undue bias. Therefore, a key question arises: is this bias significant enough to skew results, or is it negligible? In contrast to the previous approach, our cross-sectional study aims to assess the performance of two different estimation methods for the regression coefficient in a logistic regression model. Unlike the standard method, a Maximum Likelihood Estimation (MLE) model integrates possible misclassification, more thoroughly examining a possible causal relationship.

Suppose we are interested in a binary exposure status Z and binary outcome Y . We modelled the relationship between exposure and outcome by the logistic regression model (**Equation 1**), where α denotes the log-odds of experiencing an adverse event when the predictor variable is zero, equalling the prevalence of the outcome in the unvaccinated group. The regression coefficient β signifies the change in log odds of experiencing an adverse event for an unvaccinated person compared to a vaccinated person.

$$\text{logit} [\Pr (Y = 1| Z)] = \alpha + \beta Z$$

(Equation 1)

Our investigation primarily focuses on the relationship between COVID-19 vaccine status and the likelihood of experiencing an AESI; our model consists of a binary outcome event, Y , representing AESI occurrence. The binary predictor variable, Z , signifies the vaccine status; hence our naïve model considers four possibilities (**Table 1**).

Exposure (Z)	Outcome (Y)	
	Yes	No
Yes	$[\Pr(Y = 1 Z = 1)]$	$[\Pr(Y = 0 Z = 1)]$
No	$[\Pr(Y = 1 Z = 0)]$	$[\Pr(Y = 0 Z = 0)]$

Table 1. This contingency table shows the probabilities between the outcome and the exposure.

We define the observed outcome as Y^* to account for possible misclassification. Four possibilities exist for observing Y^* depending on the true outcome Y (**Table 2**). Therefore, the relationship between the observed outcome Y^* and the true observed outcome Y can again be modelled using the logistic regression model (**Equation 2**), in which τ represents the log-odds of observing an adverse event when truly no adverse event occurred, while λ indicates the change in log-odds from observing an adverse event to not observing an adverse event, similar to **Equation 1**.

Observed (Y*)	True (Y)	
	Yes	No
Yes	$[Pr(Y^* = 1 Y = 1)]$	$[Pr(Y^* = 1 Y = 0)]$
No	$[Pr(Y^* = 0 Y = 1)]$	$[Pr(Y^* = 0 Y = 0)]$

Table 2. This contingency table shows the probabilities between the true and observed outcomes.

$$\text{logit}[\Pr(Y^* = 1 | Y)] = \tau + \lambda Y$$

(Equation 2)

We consider a naïve model which assumes Y^* is Y , and therefore the estimation of the log of odds ratio of β was done using the generalised linear model with a binomial distribution. In order to account for misclassification, we model the marginal probability Y^* as follows:

$$\Pr(Y^* = 1) = \Pr(Y^* = 1 | Y = 1)\Pr(Y = 1) + \Pr(Y^* = 1|Y = 0)[1 - \Pr(Y = 1)]$$

(Equation 3)

Equation 3 essentially represents the probability of Y^* equaling one given the conditional probabilities of Y^* given Y and the overall probability of Y being 1. To accommodate this in a distributional framework, we leverage the Bernoulli distribution, which is suitable for binary outcomes like Y^* . The general form of a Bernoulli distribution is:

$$\Pr(X = x) = p^x * (1 - p)^{(1 - x)}, \text{ for } x \in \{0,1\}$$

(Equation 4)

In this context, we can express the probability distribution of Y^* as:

$$\Pr(Y^* = y) = p^y(1 - p)^{(1-y)}, \text{ for } y \in \{0,1\}$$

(Equation 5)

In Equation 5, p is the probability of Y^* being 1, as obtained from Equation 3. With these definitions in place, we can express the likelihood contribution of each subject as:

$$L_i = \Pr(Y^*_i)$$

(Equation 6)

We will compare the estimated log of odds ratio and the relative risk obtained by the naïve and the MLE method through a simulation study in which the parameters can be found in the next section. The relative risk is calculated as the ratio of the probability of being vaccinated to the probability of being unvaccinated; this gives the estimated risk of the adverse event occurring for vaccinated individuals compared to unvaccinated individuals. In other words, it tells us how much more likely it is for a vaccinated individual. In order to calculate the relative risk, we use the following formula:

$$RR = \frac{[\Pr(Y = 1|Z = 0)]}{[\Pr(Y = 0|Z = 0)]}$$

(Equation 7)

Methodology for Simulation study

The literature review provided supporting information for the conditions under which each AESI may exist. We calculated minimum and maximum sensitivity and specificity values for each AESI using the range of Positive Predictive Values (PPVs) with the formula below. When the literature review provided a PPV range, we used it directly – but for single PPV reports, we introduced a margin of error by adding a range of plus-minus 5 per cent to the given PPV, a common practice in statistics when there is uncertainty about the true value of a parameter. This information is presented in a comprehensive matrix that illustrates the possible combinations for each AESI (**Table 4**). We also present the minimal and maximal sensitivity and specificity values derived from the range of PPVs for each AESI, which can aid in the visual assessment of figures produced by the simulation study.

$$\text{PPV} = \frac{(\text{Prevalence}) \times (\text{Sensitivity})}{(\text{Prevalence}) \times (\text{Sensitivity}) + (1 - \text{Specificity}) \times (1 - \text{Prevalence})}$$

(Equation 8)

In order to assess the performance of the two estimators, we will conduct a simulation study. This study will generate data with a fixed vaccination rate, varying sample sizes, regression coefficients, and misclassification rates. The simulation study investigated the performance of various statistical methods in estimating the association between vaccination status and the occurrence of an AESI. Specifically, the study focused on the accuracy and efficiency by comparing bias and Mean Squared Error (MSE) under the various simulation scenarios (**Table 3**). The bias and MSE of an estimated parameter θ are calculated as follows:

$$\text{Bias} = E[\hat{\theta}] - \theta$$

(Equation 9)

$$\text{MSE} = [E(\hat{\theta} - \theta)^2]$$

(Equation 10)

The analysis was performed using R version 4.1.1 and associated packages [found in **Appendix 3**]. The starting point for the analysis involved changing the code from a list of binary data to frequency data, as it is faster to compute simulations with. The codes used were extended from a pre-existing code written by other team members [found in **Appendix 4**].

Condition	Description	Values
N	Sample Size	200, 1000, 2000, 10000
P_vac	Proportion of Vaccinated	0.7
Beta True	True Logistic Regression Coefficient of Vaccination on Event	0.1, 0.5, 0.9
Alpha True	True Logistic Regression Intercept of Vaccination on Event	-4.5, -3, -0.5
Sens	Sensitivity of Event Registration ($\Pr(Y^* = 1 Z = 1)$)	0.99, 0.9, 0.8, 0.7
Spec	Specificity of Event Registration ($\Pr(Y^* = 0 Z = 0)$)	0.99, 0.9, 0.8, 0.7

Table 3. The conditions of the simulation study.

Results

In this section, we illustrated the performance of two methods in estimating the association between outcome and exposure based on the specified conditions (**Table 3**). Specifically, we examined how different degrees of misclassification can influence our understanding of this relationship when misclassification is overlooked compared to when it is explicitly acknowledged. This approach can illuminate whether it is more advantageous to correct for misclassification or to disregard it. The primary performance measures assessed were bias and MSE on the log of odds ratio and the relative risk.

Bias Analysis for the Estimated Log of Odds

In the simulated scenario where the true beta value (representing the strength of association between exposure and outcome) was set to 0.1, both estimators resulted in low bias estimates (**Fig 1**). More specifically, the naïve estimator showed consistency across all conditions, regardless of sample size and true alpha values (which represent the prevalence). However, The MLE model, displayed more variability, with notable fluctuations at sample sizes of 1,000 and 2,000 but appeared to stabilise at 10,000. These estimates increased as specificity (the ability to correctly identify non-cases) rose from 0.7 to 0.99. These fluctuations were more pronounced when the prevalence was low (alpha = -4.5), whereas more minor changes were observed for higher prevalence values (alpha values of -3 and -0.5).

When we increased the strength of association (beta) to 0.5 (**Fig 2**) or 0.9 (**Fig 3**), the behaviour of both estimators remained similar however, an increase in bias was observed. The bias was substantial when the strength of association was high (beta = 0.9). However, regardless of the strength of association (beta), the bias tended to shrink as specificity increased from 0.7 to 0.99 for the naïve estimator. In the scenarios where specificity was high (0.99), the prevalence was high (alpha = -0.5), and the sample size was large (10,000), bias estimates approached zero for the naïve estimator but not for the MLE. There was an increase in bias as we increased beta.

Disease	Min Spec	Max Spec	Min Sens	Max Sens
Guillain Barre Syndrome	0.7	0.99	0.7	0.99
ADEM	0.99	0.99	0.7	0.8
Acute Aseptic Arthritis	0.7	0.7	0.7	0.9
Diabetes Mellitus	0.7	0.99	0.7	0.99
ITP	0.7	0.99	0.7	0.99
Microangiopathy	0.7	0.8	0.7	0.99
Heart Failure	0.99	0.99	0.7	0.7
Coronary Artery Disease	0.8	0.99	0.7	0.99
Ventricular Arrhythmia	0.7	0.99	0.7	0.99
Myocarditis	0.7	0.8	0.7	0.99
Pericarditis	1	0	1	0
DIC	1	0	1	0
Pulmonary Embolism + DVT	0.8	0.99	0.7	0.99
Stroke	0.9	0.99	0.7	0.99
Venous Thrombosis	1	0	1	0
Acute Liver Injury	0.7	0.99	0.7	0.99
Acute Kidney Failure	0.8	0.99	0.7	0.99
Generalised Convulsions and Epilepsy	0.99	0.99	0.7	0.99
Transverse myelitis	0.7	0.8	0.7	0.99
Acute Respiratory Distress Syndrome	0.7	0.9	0.7	0.99
COVID-19	1	0	1	0
Erythema Multiforme	0.7	0.99	0.7	0.99
Skin Infections	0.7	0.99	0.7	0.99
Anosmia, Dysgeusia	0.8	0.99	0.7	0.99
Anaphylaxis	0.7	0.99	0.7	0.99
MIS-C	0.99	0.99	0.7	0.99
Gestational Diabetes	0.7	0.9	0.7	0.99
Preeclampsia	0.7	0.99	0.7	0.99
Foetal growth restriction	0.9	0.99	0.8	0.99
Spontaneous abortions	0.9	0.99	0.7	0.99
Stillbirth	0.9	0.99	0.7	0.99
Preterm birth	0.7	0.8	0.7	0.99
Major congenital anomalies	0.7	0.99	0.7	0.99
Microcephaly	0.99	0.99	0.7	0.7
Neonatal death	0.9	0.99	0.7	0.99
TOPFA	0.99	0.99	0.7	0.99

Table 4: The minimal and maximal sensitivity and specificity values for each AESI based on the literature review.

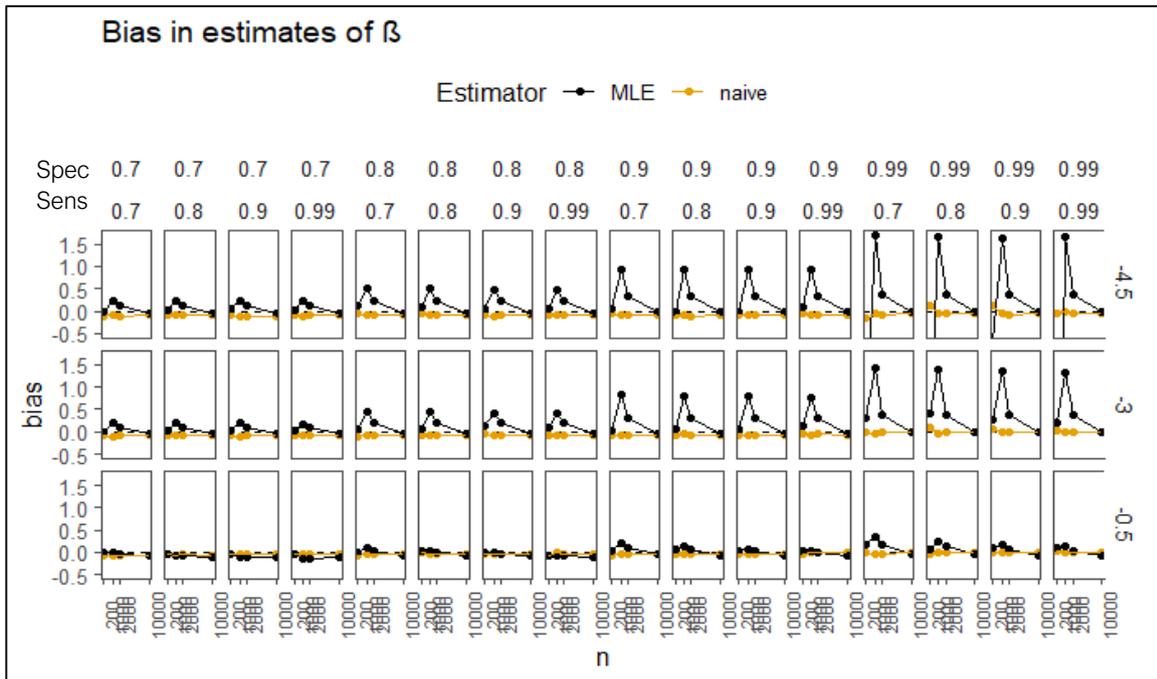


Figure 1: Assessment of Bias of the Log of Odds Ratio under the Assumption of True Beta Value Equalling 0.1

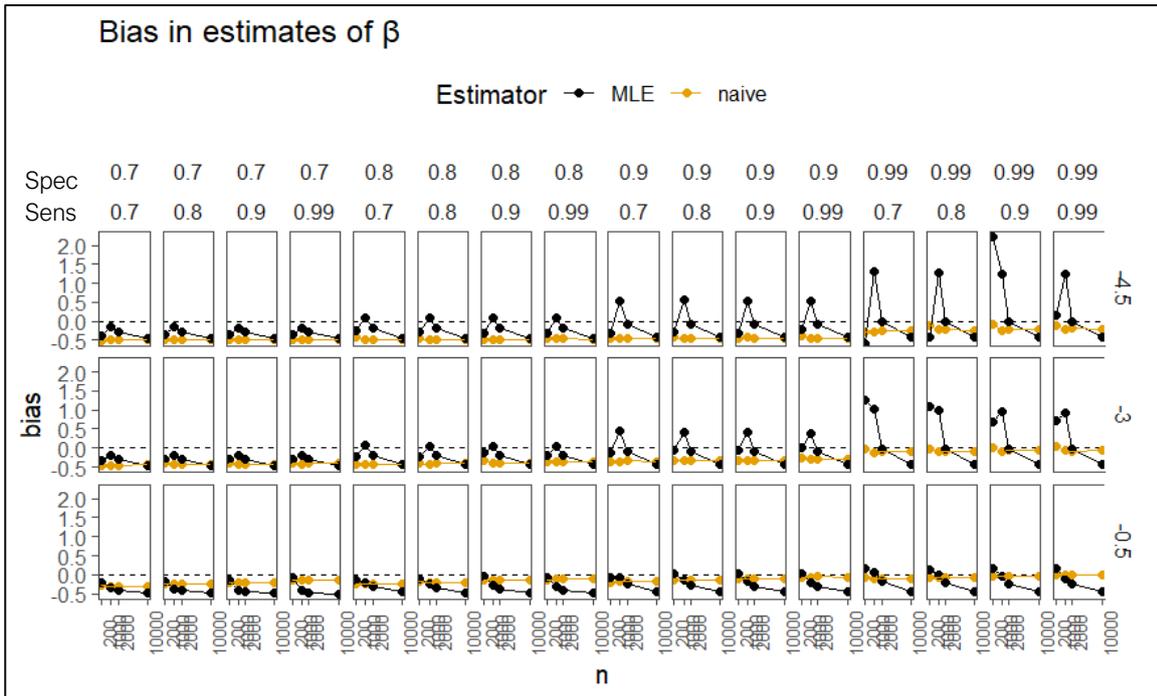


Figure 2: Assessment of Bias of the Log of Odds Ratio under the Assumption of True Beta Value Equalling 0.5

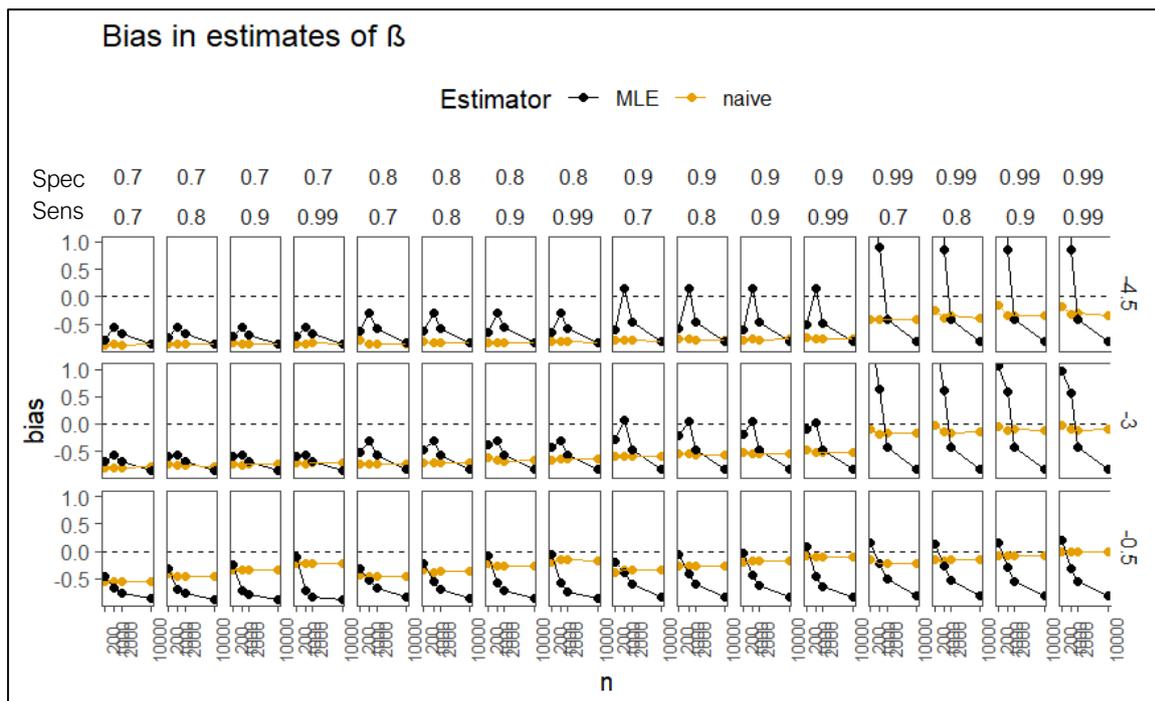


Figure 3: Assessment of Bias of the Log of Odds Ratio under the Assumption of True Beta Value Equalling 0.9

Bias Analysis for the Estimated Relative Risk

[Analysis with Figures 4, 5, and 6 found in the Appendix]

Mean Squared Error Analysis for the Estimated Log of Odds Ratio

For smaller true beta values, such as 0.1, and under conditions of lower prevalence, the MLE showed a lower MSE, with a lower MSE meaning higher accuracy, compared to the naïve estimator. This held true regardless of sample size, although both the naïve and MLE provided notably high MSE estimates (Fig 7). However, as the sample size increased to 10,000, the MLE demonstrated a significant reduction in MSE. This suggests that the MLE can be more accurate under these conditions. However, under high prevalence conditions (i.e., when alpha is set at -0.5), the MSE estimates of the two estimators alternated irregularly, sometimes surpassing and at other times falling beneath that of the naïve estimator, a phenomenon observed for all simulations when beta was 0.5 (Fig 8).

When sensitivity and specificity were both high (0.99), the MSE of the MLE and the naïve estimator performed similarly. However, a divergence was observed as the sample size increased to 10,000: under these specific conditions, the naïve estimator yielded a lower MSE, suggesting minimal misclassification. As the true beta value was increased to 0.5 and further to 0.9 (Fig 9), the results did not clearly indicate which estimator – MLE or naïve – presented lower MSEs. The resulting ambiguity emphasises the need for further exploration into the factors influencing these results.

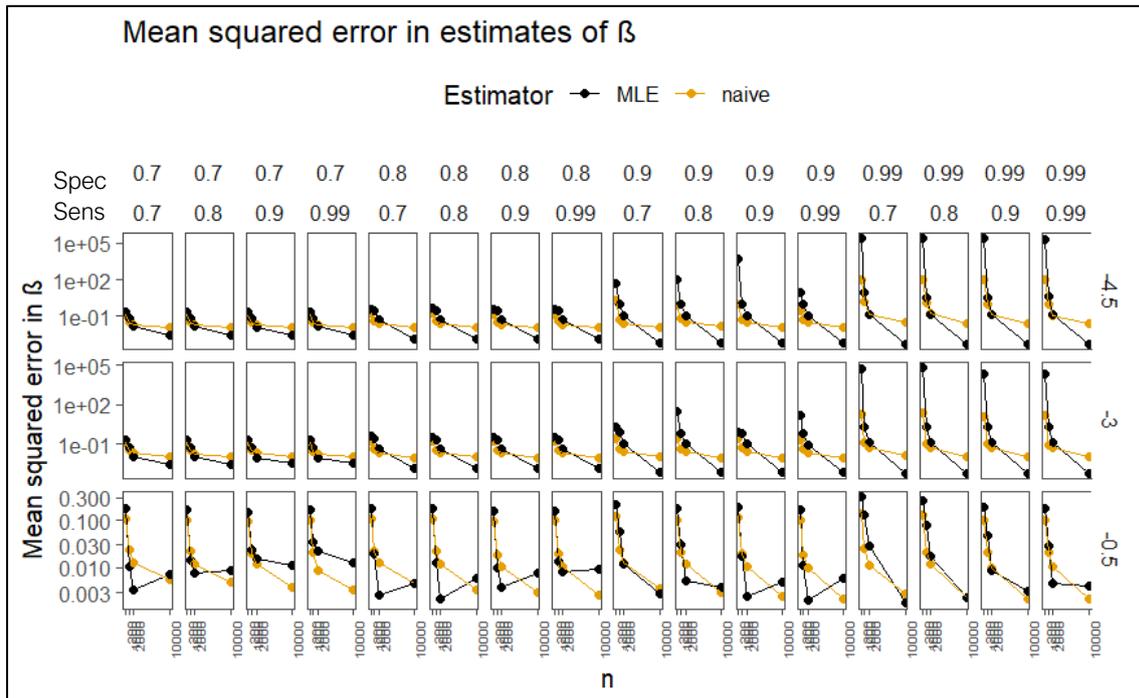


Figure 7: Assessment of MSE of the Log of Odds Ratio under the Assumption of True Beta Value Equalling 0.1.

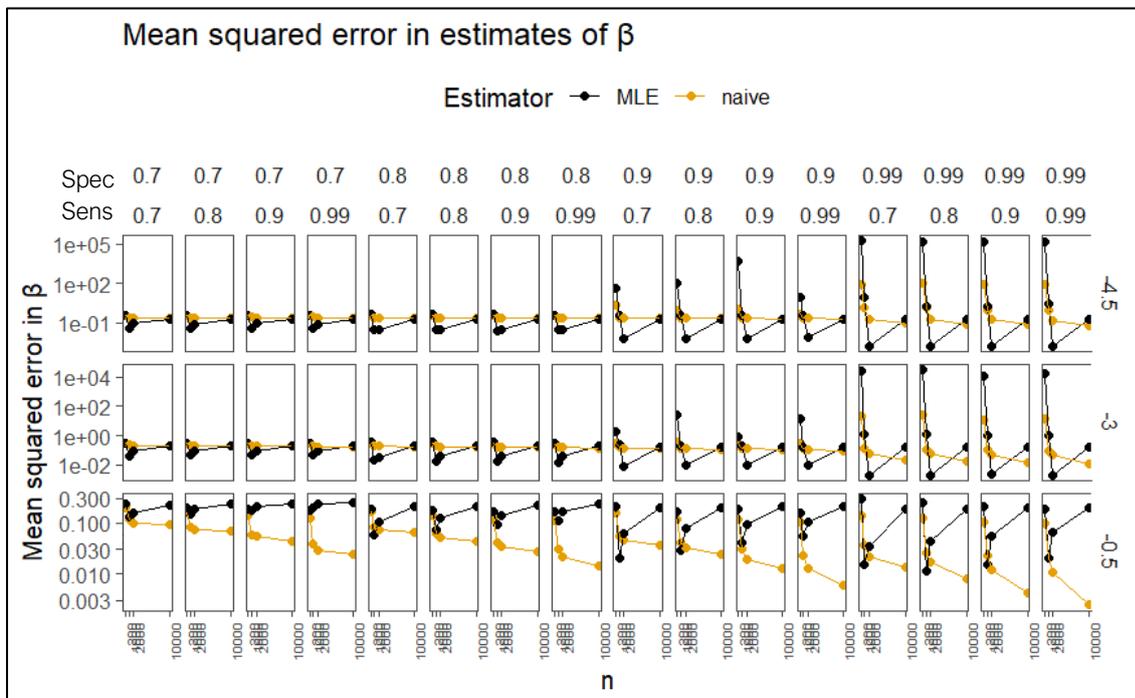


Figure 8: Assessment of MSE of the Log of Odds Ratio under the Assumption of True Beta Value Equalling 0.5.

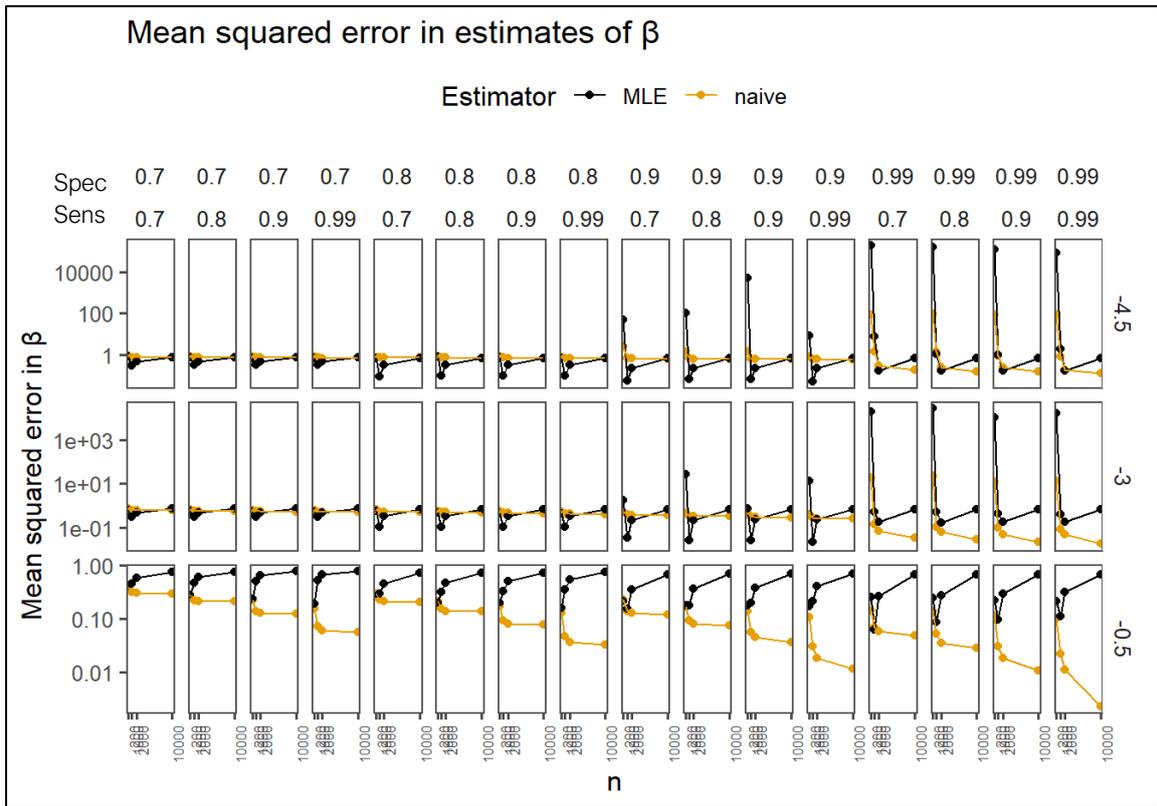


Figure 9: Assessment of MSE of the Log of Odds Ratio under the Assumption of True Beta Value Equalling 0.9.

Mean Squared Error Analysis for the Estimated Relative Risk

[Analysis with Figures 10, 11, and 12 found in the Appendix]

Discussion

Our research adds to the growing body of literature that aims to assess the impact of outcome misclassification on reporting the association of AESIs with COVID-19 vaccination. Our literature review revealed a wide range of possible sensitivity and specificity values for the pre-selected AESIs. This underlying diversity highlights the necessity for a more comprehensive understanding of these measures when researching the association of COVID-19 vaccines with AESIs.

For instance, the ICD-9 and ICD-10 codes used to classify AESIs, such as those for ADEM, Heart Failure, and Generalised Convulsions and Epilepsy, demonstrated a high minimum specificity (0.99) – signalling a high rate of true negatives. This implies that the algorithms accurately identified individuals not suffering from these AESIs, indicating that misclassification is already minimal. Conversely, the ICD codes for Acute Aseptic Arthritis, Microangiopathy, Myocarditis, and Transverse myelitis, displayed lower maximum specificity values. This fact suggests a higher risk for false negatives, an essential factor for clinicians and public health officials to consider; misclassifying these events could lead to imprecise vaccine risk assessments. Interestingly, several ICD codes for AESIs, such as Pericarditis, Disseminated Intravascular Coagulation, and Cerebral Venous Thrombosis, displayed extreme sensitivity and specificity values (1 or 0). This could be attributed to reporting errors or limited data availability, emphasising the need for consistent monitoring and data collection for accurate value estimation.

Furthermore, some codes for conditions like Guillain Barre Syndrome, Diabetes Mellitus, ITP, and coronary artery disease showed a wide range for both sensitivity and specificity, meanwhile codes for Acute Respiratory Distress Syndrome, Gestational Diabetes, Foetal Growth Restriction, and Preterm-birth exhibited narrower ranges. These disparities are particularly significant for the recognition of misclassification errors and warrant further investigation. This observation encourages the development of appropriate methods to manage misclassification in vaccine safety studies, as failure to adjust for this variability could lead to incorrect risk estimates. It is crucial to note that the broad range in sensitivity and specificity values across different AESIs' ICD codes implies that misclassification could vary significantly depending on the specific AESI and code being studied.

Furthermore, the present study underlines the importance of a holistic approach to vaccine safety. Through this approach, researchers are able to account for not just the risk of misclassification across all AESIs, but also the specific rate of the same for each event and its respective ICD code. This enhanced understanding will provide a more detailed image of the potential adverse effects related to COVID-19 immunisation and encourage informed decision-making in relation to future vaccine administration and public health policy. Our literature review lays the foundation for further research into the incorporation of misclassification rates into adverse event monitoring. In our methodology, we aimed to expose the potential bias from misclassification, which can often confound the determination of causal relationships in vaccine safety studies. The power of an RCT's causal inference can be weakened by these errors; therefore, our cross-sectional study aims to demonstrate how error and bias manifest in relation to the recognition of misclassification when it is present.

We compared a naïve estimator (which does not address misclassification) with an MLE model (which does integrate misclassification) to provide a more in-depth depiction of the association between vaccine exposure and AESIs. Both models use binary exposure status and binary outcome variables. In order to accurately portray outcome misclassification data, our MLE model was formulated to determine the marginal probability of the observed outcome based on the Bernoulli distribution. This approach enabled us to express the likelihood contribution of each subject, which thereby formed the groundwork for comparing the log of odds ratio and the relative risk obtained by our two models.

In our simulation study, we generated data which incorporated a fixed vaccination rate, varying sample sizes, regression coefficients, and misclassification rates. When simulating the bias for the log of odds ratio, we found that the naïve estimator consistently yielded lower bias estimates across all conditions when the strength of association was minimal. In contrast, the MLE model displayed a higher degree of variability, particularly at smaller sample sizes of 1,000 and 2,000. Notably, the MLE appeared to stabilise at larger sample sizes, suggesting that misclassification becomes less of an issue as the sample size increases. Furthermore, our analysis showed that bias estimates are sensitive to changes in effect size. As we increased this parameter, we observed a corresponding increase in bias for both estimators. We found that higher specificity (indicative of better identification of non-cases) helped to offset this rise in bias. Moreover, we noticed that under conditions of high specificity, high prevalence, and large sample size, bias estimates approached zero.

When examining the bias in the relative risk, we observed similar patterns as for the log of odds. The fluctuations in bias appeared to be more noticeable at lower prevalence rates with both the naïve and MLE models, demonstrating similar performance. However, as the strength of association was increased to its highest level (0.9), the bias analysis for the estimated relative risk revealed an interesting dynamic. While the naïve estimator remained relatively constant, the MLE showed an apparent decrease in bias under high prevalence conditions and high specificity. This reduction was particularly pronounced in simulations with the largest effect size, highlighting the significant impact of bias and misclassification when predicting large effect sizes. These findings underscore the potential benefit of using the MLE to effectively adjust for misclassification, potentially leading to a decrease in estimate bias.

Under low prevalence and small effect sizes, the MLE displayed higher accuracy than the naïve estimator, evidenced by lower MSE estimates in log of odds. This trend persisted across sample sizes, with MSE estimates rising for both models while sample sizes were small. However, the MLE's performance fluctuated under high prevalence and increased effect sizes, with the estimates occasionally surpassing and falling beneath the naïve estimator's MSE.

For the relative risk, the naïve and MLE showed similar estimates for the small effect size simulations. After we increased the effect size value to 0.5 and 0.9, the MLE consistently outperformed the naïve estimator, indicating better error minimisation and affirming our hypothesis. Despite our proposed method of having more negligible bias but more considerable variability, this suggests that it may not necessarily outperform a less accurate, yet more biased naïve estimate (i.e. an estimator that is near to the true value but exhibits high variability, may not necessarily outperform a naive estimate, which, although biased, demonstrates less variability).

The analysis included a computational simulation involving dataset simulation, defining conditions, and the use of various estimators. Using a random seed ensured reproducibility in results, and implementing a large number of simulations guarantees a comprehensive examination of potential real-world scenarios. Moreover, the computation and visualisation of error metrics provide an extensive understanding of the model's performance. Nevertheless, the

complexity of the MLE method and its implementation has limitations, because it is computationally expensive, time-consuming and susceptible to potential issues, such as non-convergence. While this adds to the robustness of our research, it also increases the resources required. Overall, the benefits of this thorough approach outweigh the disadvantages, contributing to the study's impact on the field of epidemiology.

Conclusion

The insights from our literature review emphasise the importance of a holistic approach to studying vaccine safety. We aimed to expose the potential bias and error resulting from data misclassification which can confound the determination of causal relationships in vaccine safety studies. These errors can deteriorate the power of causal inferences; thus, our cross-sectional study demonstrated how error and bias are exhibited through misclassification when present.

The MLE and naïve estimators' varied results regarding MSE necessitate a careful and critical interpretation. These varied findings highlight the need to improve our understanding of the factors shaping these results. Essentially, our study aims to investigate how incorporating misclassification into the modelling of outcome-exposure relationships can approximate the true mechanisms of the association, as determined through bias and MSE. The importance of this research lies in its emphasis on adopting robust methods to address misclassification and recognition of the unique characteristics pertinent to specific AESI and their ICD codes.

Our work contributes to the gap in research concerning the safety of COVID-19 vaccines and pinpoints critical areas for future exploration in this field. Our observations show that our proposed method of incorporating misclassification can sometimes be beneficial, but under certain conditions, it may not be necessary.

REFERENCES

- Aldawas, A., Ishfaq, M., 2022. COVID-19: Multisystem Inflammatory Syndrome in Children (MIS-C). *Cureus* 14, e21064. <https://doi.org/10.7759/cureus.21064>
- Andrade, S.E., Shinde, M., Moore Simas, T.A., Bird, S.T., Bohn, J., Haynes, K., Taylor, L.G., Luring, J.R., Longley, E., McMahonill-Walraven, C.N., Trinacty, C.M., Saphirak, C., Delude, C., DeLuccia, S., Zhang, T., Cole, D.V., DiNunzio, N., Gertz, A., Fazio-Eynullayeva, E., Stojanovic, D., 2021. Validation of an ICD-10-based algorithm to identify stillbirth in the Sentinel System. *Pharmacoepidemiology and Drug Safety* 30, 1175–1183. <https://doi.org/10.1002/pds.5300>
- Bezin, J., Girodet, P.-O., Rabelomanana, S., Touya, M., Ferreira, P., Gilleron, V., Robinson, P., Moore, N., Pariente, A., 2015. Choice of ICD-10 codes for the identification of acute coronary syndrome in the French hospitalisation database. *Fundam Clin Pharmacol* 29, 586–591. <https://doi.org/10.1111/fcp.12143>
- Bhat, A.G., White, K., Gobeil, K., Lagu, T., Lindenauer, P.K., Pack, Q.R., 2020. Utility of ICD Codes for Stress Cardiomyopathy in Hospital Administrative Databases: What Do They Signify? *Journal of Hospital Medicine* 15, 160–163. <https://doi.org/10.12788/jhm.3344>
- Boesen, M.S., Magyari, M., Born, A.P., Thygesen, L.C., 2018. Pediatric acquired demyelinating syndromes: a nationwide validation study of the Danish National Patient Register. *Clin Epidemiol* 10, 391–399. <https://doi.org/10.2147/CLEP.S156997>
- Bok, K., Sitar, S., Graham, B.S., Mascola, J.R., 2021. Accelerated COVID-19 vaccine development: milestones, lessons, and prospects. *Immunity* 54, 1636–1651. <https://doi.org/10.1016/j.immuni.2021.07.017>
- Burwen, D.R., Sandhu, S.K., MaCurdy, T.E., Kelman, J.A., Gibbs, J.M., Garcia, B., Markatou, M., Forshee, R.A., Izurieta, HS, Ball, R., 2012. Surveillance for Guillain–Barré Syndrome After Influenza Vaccination Among the Medicare Population, 2009–2010. *Am J Public Health* 102, 1921–1927. <https://doi.org/10.2105/AJPH.2011.300510>
- Chomistek, A.K., Phiri, K., Doherty, M.C., Calderbank, J.F., Chiuve, S.E., McIlroy, B.H., Snabes, M.C., Enger, C., Seeger, J.D., 2023. Development and Validation of ICD-10-based Algorithms for Date of Last Menstrual Period, Pregnancy Outcomes, and Infant Outcomes. *Drug Saf* 46, 209–222. <https://doi.org/10.1007/s40264-022-01261-5>
- de Sordi, D., Kappen, S., Otto-Sobotka, F., Kulschewski, A., Weyland, A., Gutierrez, L., Fortuny, J., Reinold, J., Schink, T., Timmer, A., 2021. Validity of hospital ICD-10-GM codes to identify anaphylaxis. *Pharmacoepidemiology and Drug Safety* 30, 1643–1652. <https://doi.org/10.1002/pds.5348>
- Donovan, B.M., Breheny, P.J., Robinson, J.G., Baer, R.J., Saftlas, A.F., Bao, W., Greiner, A.L., Carter, K.D., Oltman, S.P., Rand, L., Jelliffe-Pawlowski, L.L., Ryckman, K.K., 2019. Development and validation of a clinical model for preconception and early pregnancy risk prediction of gestational diabetes mellitus in nulliparous women. *PLoS ONE* 14, e0215173. <https://doi.org/10.1371/journal.pone.0215173>
- Eisenberg, D.F., Daniel, G.W., Jones, J.K., Goehring Jr, E.L., Wahl, P.M., Winters, P., Levin, J., Bohn, R.L., 2012. Validation of a claims-based diagnostic code for Stevens–Johnson syndrome in a commercially insured population. *Pharmacoepidemiology and Drug Safety* 21, 760–764. <https://doi.org/10.1002/pds.3276>
- Funch, D., Holick, C., Velentgas, P., Clifford, R., Wahl, P.M., McMahonill-Walraven, C., Gladowski, P., Platt, R., Amato, A., Chan, K.A., 2013. Algorithms for identification of Guillain–Barré Syndrome among adolescents in claims databases. *Vaccine* 31, 2075–2079. <https://doi.org/10.1016/j.vaccine.2013.02.009>

- Galdarossa, M., Vianello, F., Tezza, F., Allemand, E., Treleani, M., Scarparo, P., Fabris, F., 2012. Epidemiology of primary and secondary thrombocytopenia: first analysis of an administrative database in a major Italian institution. *Blood Coagulation & Fibrinolysis* 23, 271–277. <https://doi.org/10.1097/MBC.0b013e328351882d>
- Garne, E., Urhoj, S.K., Bakker, M., Gissler, M., Given, J., Heino, A., Limb, E., Loane, M., de Walle, H., Morris, J., 2023. The quality and the accuracy of codes for terminations of pregnancy for fetal anomalies recorded in hospital databases in three countries in northern Europe. *Birth Defects Research* 115, 405–412. <https://doi.org/10.1002/bdr2.2133>
- Geller, S.E., Ahmed, S., Brown, M.L., Cox, S.M., Rosenberg, D., Kilpatrick, S.J., 2004. International classification of diseases–9th revision coding for preeclampsia: how accurate is it? *American Journal of Obstetrics and Gynecology* 190, 1629–1633. <https://doi.org/10.1016/j.ajog.2004.03.061>
- Gillmeyer, K.R., Lee, M.-M., Link, A.P., Klings, E.S., Rinne, S.T., Wiener, R.S., 2019. Accuracy of Algorithms to Identify Pulmonary Arterial Hypertension in Administrative Data. *Chest* 155, 680–688. <https://doi.org/10.1016/j.chest.2018.11.004>
- Gmuca, S., Hardy, D.I., Narula, S., Stoll, S., Harris, J., Zhao, Y., Xiao, R., Weiss, P.F., Waldman, A.T., Gerber, J.S., 2020. Validation of claims-based diagnoses of adult and pediatric neuromyelitis optica spectrum disorder and variations in diagnostic evaluation and treatment initiation. *Multiple Sclerosis and Related Disorders* 37, 101488. <https://doi.org/10.1016/j.msard.2019.101488>
- Graham, B.S., 2020. Rapid COVID-19 vaccine development. *Science* 368, 945–946. <https://doi.org/10.1126/science.abb8923>
- Gray, K., Cameron, S., McKenzie, K., Miller, M., Odoardi, N., Tijssen, J.A., 2022. Validation of ICD-10 codes for the identification of paediatric out-of-hospital cardiac arrest patients. *Resuscitation* 171, 73–79. <https://doi.org/10.1016/j.resuscitation.2021.12.017>
- Handley, J.D., Emsley, H.C., 2020. Validation of ICD-10 codes shows intracranial venous thrombosis incidence to be higher than previously reported. *Health Inf Manag* 49, 58–61. <https://doi.org/10.1177/1833358318819105>
- Harrold, L.R., Saag, K.G., Yood, R.A., Mikuls, T.R., Andrade, S.E., Fouayzi, H., Davis, J., Chan, K.A., Raebel, M.A., Von Worley, A., Platt, R., 2007. Validity of gout diagnoses in administrative data. *Arthritis Rheum* 57, 103–108. <https://doi.org/10.1002/art.22474>
- Heden, K., Jensen, A., Farkas, D., Nørgaard, M., 2009. Validity of a procedure to identify patients with chronic idiopathic thrombocytopenic purpura in the Danish National Registry of Patients. *Clinical epidemiology* 1, 7–10.
- International Classification of Diseases (ICD) [WWW Document], n.d. URL <https://www.who.int/standards/classifications/classification-of-diseases> (accessed 6.12.23).
- Jetté, N., Reid, A.Y., Quan, H., Hill, M.D., Wiebe, S., 2010. How accurate is ICD coding for epilepsy? *Epilepsia* 51, 62–69. <https://doi.org/10.1111/j.1528-1167.2009.02201.x>
- Kanitakis, J., Lesort, C., Danset, M., Jullien, D., 2020. Chilblain-like acral lesions during the COVID-19 pandemic ("COVID toes"): Histologic, immunofluorescence, and immunohistochemical study of 17 cases. *J Am Acad Dermatol* 83, 870–875. <https://doi.org/10.1016/j.jaad.2020.05.145>
- Khalili, D., Mosavi-Jarrahi, A., Eskandari, F., Mousavi-Jarrahi, Y., Hadaegh, F., Mohagheghi, M., Azizi, F., 2012. Evaluation of Cause of Deaths' Validity Using Outcome Measures from a Prospective, Population Based Cohort Study in Tehran, Iran. *PLOS ONE* 7, e31427. <https://doi.org/10.1371/journal.pone.0031427>
- Khokhar, B., Jette, N., Metcalfe, A., Cunningham, C.T., Quan, H., Kaplan, G.G., Butalia, S., Rabi, D., 2016. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open* 6, e009952. <https://doi.org/10.1136/bmjopen-2015-009952>

- Kluberg, S.A., Hou, L., Dutcher, S.K., Billings, M., Kit, B., Toh, S., Dublin, S., Haynes, K., Kline, A., Maiyani, M., Pawloski, P.A., Watson, E.S., Cocoros, N.M., 2022. Validation of diagnosis codes to identify hospitalised COVID-19 patients in health care claims data. *Pharmacoepidemiology and Drug Safety* 31, 476–480. <https://doi.org/10.1002/pds.5401>
- Levine, P.J., Elman, M.R., Kullar, R., Townes, J.M., Bearden, D.T., Vilches-Tran, R., McClellan, I., McGregor, J.C., 2013. Use of electronic health record data to identify skin and soft tissue infections in primary care settings: a validation study. *BMC Infect Dis* 13, 171. <https://doi.org/10.1186/1471-2334-13-171>
- Lo Re, V., Haynes, K., Goldberg, D., Forde, K.A., Carbonari, D.M., Leidl, K.B.F., Hennessy, S., Reddy, K.R., Pawloski, P.A., Daniel, G.W., Cheetham, T.C., Iyer, A., Coughlin, K.O., Toh, S., Boudreau, D.M., Selvam, N., Cooper, W.O., Selvan, M.S., VanWormer, J.J., Avigan, M.I., Houstoun, M., Zornberg, G.L., Racoosin, J.A., Shoaibi, A., 2013. Validity of Diagnostic Codes to Identify Cases of Severe Acute Liver Injury in the US Food and Drug Administration's Mini-Sentinel Distributed Database. *Pharmacoepidemiol Drug Saf* 22, 861–872. <https://doi.org/10.1002/pds.3470>
- Lynch, K., Viernes, B., Gatsby, E., DuVall, S., Jones, B., Box, T., Kreisler, C., Jones, M., 2021. Positive Predictive Value of COVID-19 ICD-10 Diagnosis Codes Across Calendar Time and Clinical Setting. *Clinical Epidemiology Volume* 13, 1011–1018. <https://doi.org/10.2147/CLEP.S335621>
- McCormick, N., Bhole, V., Lacaille, D., Avina-Zubieta, J.A., 2015. Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review. *PLoS One* 10, e0135834. <https://doi.org/10.1371/journal.pone.0135834>
- McDowell Torres, D., Stevens, R.D., Gurakar, A., 2010. Acute Liver Failure. *Gastroenterol Hepatol (N Y)* 6, 444–450.
- Mesfin, Y.M., Cheng, A.C., Tran, A.H.L., Buttery, J., 2019. Positive predictive value of ICD-10 codes to detect anaphylaxis due to vaccination: A validation study. *Pharmacoepidemiol Drug Saf* 28, 1353–1360. <https://doi.org/10.1002/pds.4877>
- Mullooly, J.P., Donahue, J.G., DeStefano, F., Baggs, J., Eriksen, E., for the VSD Data Quality Working Group, 2008. Predictive Value of ICD-9 Codes Used in Vaccine Safety Research. *Methods Inf Med* 47, 328–335. <https://doi.org/10.3414/ME0500>
- Myers, R.P., Leung, Y., Shaheen, A.A.M., Li, B., 2007. Validation of ICD-9/ICD-10 coding algorithms for the identification of patients with acetaminophen overdose and hepatotoxicity using administrative data. *BMC Health Services Research* 7, 159. <https://doi.org/10.1186/1472-6963-7-159>
- Öhman, L., Johansson, M., Jansson, J.-H., Lind, M., Johansson, L., 2018. Positive predictive value and misclassification of diagnosis of pulmonary embolism and deep vein thrombosis in Swedish patient registries. *Clin Epidemiol* 10, 1215–1221. <https://doi.org/10.2147/CLEP.S177058>
- Patel, M.M., Jackson, M.L., Ferdinands, J., 2020. Postlicensure Evaluation of COVID-19 Vaccines. *JAMA* 324, 1939–1940. <https://doi.org/10.1001/jama.2020.19328>
- Phiri, K., 2015. Accuracy of ICD-9 coding to identify small for gestational age newborns - PMC [WWW Document]. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4533984/> (accessed 3.3.23).
- Pocobelli, G., Ichikawa, L., Yu, O., Green, B.B., Meyers, K., Gray, R., Shea, M., Chubak, J., 2022. Validation of international classification of diseases, tenth revision, clinical modification diagnosis codes for heart failure subtypes. *Pharmacoepidemiology and Drug Safety* 31, 992–997. <https://doi.org/10.1002/pds.5489>
- Salemi, J.L., Rutkowski, R.E., Tanner, J.P., Matas, J.L., Kirby, R.S., 2018. Identifying Algorithms to Improve the Accuracy of Unverified Diagnosis Codes for Birth Defects. *Public Health Rep* 133, 303–310. <https://doi.org/10.1177/0033354918763168>
- Schneider, G., Kachroo, S., Jones, N., Crean, S., Rotella, P., Avetisyan, R., Reynolds, M.W., 2012. A systematic review of validated methods for identifying erythema multiforme major/minor/not otherwise

- specified, Stevens-Johnson Syndrome, or toxic epidermal necrolysis using administrative and claims data. *Pharmacoepidemiol Drug Saf* 21 Suppl 1, 236–239. <https://doi.org/10.1002/pds.2331>
- Shui, I.M., Shi, P., Dutta-Linn, M.M., Weintraub, E.S., Hambidge, S.J., Nordin, J.D., Lieu, T.A., 2009. Predictive value of seizure ICD-9 codes for vaccine safety research. *Vaccine* 27, 5307–5312. <https://doi.org/10.1016/j.vaccine.2009.06.092>
- Sigakis, M.J.G., Leffert, L.R., Mirzakhani, H., Sharawi, N., Rajala, B., Callaghan, W.M., Kuklina, E.V., Creanga, A.A., Mhyre, J.M., Bateman, B.T., 2016. The Validity of Discharge Billing Codes Reflecting Severe Maternal Morbidity. *Anesth Analg* 123, 731–738. <https://doi.org/10.1213/ANE.0000000000001436>
- Tanpowpong, P., Lertudomphonwanit, C., Phuapradit, P., Treepongkaruna, S., 2021. Value of the International Classification of Diseases code for identifying children with biliary atresia. *Clin Exp Pediatr* 64, 80–85. <https://doi.org/10.3345/cep.2020.00423>
- Thomsen, G.E., Morris, A.H., 1995. Incidence of the adult respiratory distress syndrome in the state of Utah. *Am J Respir Crit Care Med* 152, 965–971. <https://doi.org/10.1164/ajrccm.152.3.7663811>
- Wada, T., Shiraishi, A., Gando, S., Yamakawa, K., Fujishima, S., Saitoh, D., Kushimoto, S., Ogura, H., Abe, T., Mayumi, T., Sasaki, J., Kotani, J., Takeyama, N., Tsuruta, R., Takuma, K., Yamashita, N., Shiraishi, S.-I., Ikeda, H., Shiino, Y., Tarui, T., Nakada, T.-A., Hifumi, T., Okamoto, K., Sakamoto, Y., Hagiwara, A., Masuno, T., Ueyama, M., Fujimi, S., Umemura, Y., Otomo, Y., 2021. Disseminated intravascular coagulation immediately after trauma predicts a poor prognosis in severely injured patients. *Sci Rep* 11, 11031. <https://doi.org/10.1038/s41598-021-90492-0>
- Waikar, S.S., Wald, R., Chertow, G.M., Curhan, G.C., Winkelmayer, W.C., Liangos, O., Sosa, M.-A., Jaber, B.L., 2006. Validity of International Classification of Diseases, Ninth Revision, Clinical Modification Codes for Acute Renal Failure. *JASN* 17, 1688–1694. <https://doi.org/10.1681/ASN.2006010073>
- Watson, G., Dodd, C., Munoz, F.M., Eckert, L.O., Jones, C.E., Buttery, J.P., Yildirim, I.B., Kachikis, A., Heath, P.T., Schlaudecker, E.P., Bond, N.H., Santarcangelo, P.L., Wilcox, C.R., Bellamy, K., Elmontser, M., Sienas, L., Simon, R., Khalil, A., Townsend, R., Sturkenboom, M., Black, S., 2021. Applicability of the GAIA Maternal and Neonatal Outcome Case Definitions for the Evaluation of Adverse Events Following Vaccination in Pregnancy in High-income Countries. *The Pediatric Infectious Disease Journal* 40, 1127–1134. <https://doi.org/10.1097/INF.0000000000003261>
- Ye, Y., Larrat, E.P., Caffrey, A.R., 2018. Algorithms used to identify ventricular arrhythmias and sudden cardiac death in retrospective studies: a systematic literature review. *Ther Adv Cardiovasc Dis* 12, 39–51. <https://doi.org/10.1177/1753944717745493>
- Zayet, S., Klopfenstein, T., Mercier, J., Kadiane-Oussou, N.J., Lan Cheong Wah, L., Royer, P.-Y., Toko, L., Gendrin, V., 2021. Contribution of anosmia and dysgeusia for diagnostic of COVID-19 in outpatients. *Infection* 49, 361–365. <https://doi.org/10.1007/s15010-020-01442-3>
- Zgibor, J.C., Orchard, T.J., Saul, M., Piatt, G., Ruppert, K., Stewart, A., Siminerio, L.M., 2007. Developing and validating a diabetes database in a large health system. *Diabetes Research and Clinical Practice* 75, 313–319. <https://doi.org/10.1016/j.diabres.2006.07.007>

APPENDICES

Appendix 1: Literature Review Table Summary

Disease	ICD Version	ICD Code	PPV Values	Reference
Guillain Barre Syndrome	ICD-9	357.0	12%-16%	Yih et al. (2012)
			29%	Funch et al. (2013)
			30-71%	Shui et al. (2011)
			82%	Burwen et al. (2012)
Acute Disseminated Encephalomyelitis	ICD-10	G04.0, G04.8	44%	Boesen et al. (2018)
Acute Aseptic Arthritis	ICD-9	274	61%	Harrold et al. (2007)
Diabetes Mellitus	ICD-9	250.x	94%-97%	Zgibor et al. (2007)
			54%-80%	Kohkar et al. (2016)
Idiopathic Thrombocytopenic Purpura (ITP)	ICD-10	D69.3	93%-95.8%	Heden et al. (2009)
	ICD-9	287.31	65%-85%	
Microangiopathy	ICD-9/ICD-10	448.0/I78	70%	Gillmeyer et al. (2021)
Heart Failure	ICD-10	I50.20-I50.23	41.4%	Pocobelli et al. (2022)
Coronary Artery Disease	ICD-10	I20.0, I21, I24	66.7%-100.0%	Bezin et al (2015)
Ventricular Arrhythmia	ICD-9	427.1, 427.4, 427.5	50-93%	Ye et al (2018)
Myocarditis	ICD-9, ICD-10	422, I40/I41	64%	Sundboll et al (2016)
Pericarditis	ICD-9, ICD-10	391.09, 393, 420, 423, I30-132	92%	Sundboll et al (2016)
Disseminated Intravascular Coagulation	ICD-10	D65, D68.4	89.7%-90.3%	Wada et al (2014), Okamoto et al. (2019)
Pulmonary Embolism + DVT	ICD-8/9/10		68.2-90.3%	Öhman et al (2018)
Stroke	ICD-9/10	430/I60, 431/I61, 434/I63, 436/I64	82%	McCormick et al (2015)
Cerebral Venous Thrombosis	ICD-10	G08.X, O22.5, I67.6, I63.6, O87.3.	92.3%	Handley and Emsley (2020)
Acute Liver Injury	ICD-9	570, 572.2, 572.4, 572.8, 573.3 and 570	47.1-60%	Lo Re (2013)

Acute Kidney Failure	ICD-9	584.5, 584.6, 584.7, 584.8, or 584.9.	80.2%.	Waikar et al (2006)
Generalised Convulsions and Epilepsy	ICD-9	345.x	98.9%	Jette et al (2008)
Transverse myelitis	ICD-10	G37.3	64%	Boesen et al (2018)
Acute Respiratory Distress Syndrome	ICD-9	518.5, 518.81, 518.82	7%	Thomsen & Morris (1995)
COVID-19	ICD-10	U07.1	90.0-93.0%	Kadri et al (2020), Kluberg et al (2022), Bhatt et al (2021)
Erythema Multiforme	ICD-9	695.1x	2.00%-59.6%	Eisenberg et al (2012); Schneider et al (2012)
Skin and Soft Tissue Infections	ICD-9	N/A	52.6%-91.5%	Levine et al (2013)
Anosmia, Dysgeusia	ICD-10	N/A10	77%	Zayet et al (N/A)
Anaphylaxis	ICD-10	T80.5	95.7%	Mesfin et al (2019)
		T78.2, T88.6, T80.5, T78.4, T88.7, Y57.9	12.9%-80.6%	De Sordi et al. (2021)
Multisystem Inflammatory Syndrome in Children (MIS-C)	ICD-10	U07.1	96.3%	Blatz et al (2021)
Sudden Death	ICD-9/ICD-10	798/R99, R95, R96.0, R96.1	N/A	N/A
Gestational Diabetes	ICD-9	648.8	11.6%	Donovan et al, 2019
Preeclampsia	ICD-9	642.4	45.3%	Geller et al. (2004)
	ICD-10	O14	81.2%	Watson et al. (2021)
Maternal death	ICD-9	-	-	Sigakis et al
Foetal growth restriction	ICD-9	656.5x	86.8%	Phiri et al, 2016
Spontaneous abortions	ICD-10	O03	84.7%	Chomistek et al, 2023
Stillbirth	ICD-10	Z37.1, Z37.3, Z37.4, Z37.6X, Z37.7, P95 / O36.4XXX, O31.0XXX	82.5%	Andrade et al, 2021
Preterm birth	ICD-10	O60	56.8%	Watson et al. (2021)
		P07	76.4%	
Major congenital anomalies	ICD-10	-	44.0% - 67.8%	Chomistek et al. (2023)
Microcephaly	ICD-10	Q02	40%	Watson et al. (2021)

Neonatal death	ICD-10	-	81.4%	Gray et al. (2022)
Termination of Pregnancy for Foetal Anomaly (TOPFA)	-	-	94.2%	Salemi et al. (2018)
Table 5. Table showing the specific AESI, ICD version, respective code, reported PPV, and source.				

Appendix 2: RStudio Script for the Calculation of Minimum and Maximum Sensitivity

```
# Define the PPV calculation function
calculate_ppv <- function(sens, spec, prevalence) {
  return(sens * prevalence / (sens * prevalence + (1 - spec) * (1 - prevalence)))
}
# Define the prevalence calculation function
calculate_prevalence <- function(alpha) {
  return(exp(alpha) / (1 + exp(alpha)))
}
# Define the range of spec and sens values
spec_sens_range <- c(0.7, 0.8, 0.9, 0.99)
# Define the alpha values
alpha_values <- c(-4.5, -3.0, -0.5)
# Initialise a dataframe to store the results
results <- data.frame()
# Define the diseases and their PPV ranges
diseases <- data.frame(
  Name = c("Name of each AESI in a list...
),
  PPV_lower = c(Lower PPV limit of each AESI in order,
),
  PPV_upper = c( )
)
# Loop over diseases
for (i in 1:nrow(diseases)) {
  # Loop over alpha values to calculate prevalence
  for(alpha in alpha_values){
    prevalence <- calculate_prevalence(alpha)
    # Loop over spec and sens values
    for(spec in spec_sens_range){
      for(sens in spec_sens_range){
        # Calculate PPV
        ppv <- calculate_ppv(sens, spec, prevalence)
        # Check if PPV falls within the reported range for the disease
        if(ppv >= diseases$PPV_lower[i] & ppv <= diseases$PPV_upper[i]){
          results <- rbind(results, c(diseases$Name[i], alpha, prevalence,
spec, sens, ppv))
        }
      }
    }
  }
}
# Name the columns of the results dataframe
names(results) <- c("Disease", "Alpha", "Prevalence", "Specificity",
"Sensitivity", "PPV")
# Print the results
print(results)
write.excel <- function(x,row.names=FALSE,col.names=TRUE,...) {
write.table(x,"clipboard",sep="\t",row.names=row.names,col.names=col.names,...
)
}
write.excel(results)
# Initialise a dataframe to store the results
results <- data.frame()
# Loop over diseases
for (i in 1:nrow(diseases)) {
  # Initialize min and max spec and sens values
  min_spec <- 1
  max_spec <- 0
  min_sens <- 1
  max_sens <- 0
  # Loop over alpha values to calculate prevalence
```

```

for(alpha in alpha_values){
  prevalence <- calculate_prevalence(alpha)
  # Loop over spec and sens values
  for(spec in spec_sens_range){
    for(sens in spec_sens_range){
      # Calculate PPV
      ppv <- calculate_ppv(sens, spec, prevalence)
      # Check if PPV falls within the reported range for the disease
      if(ppv >= diseases$PPV_lower[i] & ppv <= diseases$PPV_upper[i]){
        # Update min and max spec and sens values
        min_spec <- min(min_spec, spec)
        max_spec <- max(max_spec, spec)
        min_sens <- min(min_sens, sens)
        max_sens <- max(max_sens, sens)
      }
    }
  }
  # Add the results for the disease to the results dataframe
  results <- rbind(results, c(diseases$Name[i], min_spec, max_spec, min_sens,
max_sens))
}
# Name the columns of the results dataframe
names(results) <- c("Disease", "Min_Specificity", "Max_Specificity",
"Min_Sensitivity", "Max_Sensitivity")
# Print the results
print(results)
write.excel <- function(x,row.names=FALSE,col.names=TRUE,...) {

write.table(x,"clipboard",sep="\t",row.names=row.names,col.names=col.names,...
)
}
write.excel(results)

```

Appendix 3: RStudio Packages Used in Simulation Study

Packages Used:

- tidyverse (version 1.3.1): A collection of packages for data manipulation and visualisation in R.
- maxLik (version 1.4-12): A package for maximum likelihood estimation in R.
- dplyr (version 1.0.8): A package for data manipulation in R.
- tidyr (version 1.1.4): A package for data tidying in R.

Citations:

- tidyverse: Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- maxLik: Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood estimation in R. Computational Statistics 26(3), 443-458. doi:10.1007/s00180-010-0217-1
- dplyr: Wickham et al., (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.8, <https://CRAN.R-project.org/package=dplyr>
- tidyr: Wickham et al., (2021). tidyr: Tidy Messy Data. R package version 1.1.4, <https://CRAN.R-project.org/package=tidyr>

Appendix 4: The Complete RStudio-Script for the Simulation Study

Link to Daniels code: <https://github.com/daob/vaccine-misclassification>

```
set.seed(342)
simulate_dataset <- function(condition) {
  with(condition, {
    z <- rbinom(n, 1, prob = p_vaccinated)
    eta_y_true <- alpha_true + beta_true * z
    y <- rbinom(n, 1, prob = plogis(eta_y_true))
    eta_ystar_true <- tau + lambda * y
    y_star <- rbinom(n, 1, prob = plogis(eta_ystar_true))
    dat_samp <- tibble(y_star = y_star, z = z)
    dat_samp %>% group_by(y_star, z) %>% summarise(count = n(), .groups =
'drop') %>% complete(y_star, z, fill = list(count = 0))
  })
}

set.seed(342)
nsim <- 500

conditions <- expand.grid(n = c(200, 1000, 2000, 10000), p_vaccinated = 0.70,
beta_true = c(0.1, 0.5, -0.9), alpha_true = c(-4.5, -3, -5), p_true_positive
= c(0.99, 0.9, 0.8, 0.7), p_true_negative = c(0.99, 0.9, 0.8, 0.7)) %>% tibble

conditions$tau <- with(conditions, log(1 - p_true_negative) -
log(p_true_negative))
conditions$lambda <- with(conditions, log(p_true_positive) - log(1 -
p_true_positive) - tau)

condition <- conditions

estimator_naive <- function(dat) {
  fit_glm <- suppressWarnings(glm(y_star ~ z, data = dat, weights = count,
family = binomial))
  relative_risk <- with(dat, (sum(dat$y_star * dat$count * (z == 1)) /
sum(dat$count[z == 1])) / (sum(dat$y_star * dat$count * (z == 0)) /
sum(dat$count[z == 0])))
  tibble(est_alpha_glm_naive = coef(fit_glm)['(Intercept)'], est_beta_glm_naive
= coef(fit_glm)['z'], est_rr_naive = relative_risk)
}

data.new <- simulate_dataset(condition)
cond <- condition

get_estimator_misclassification_maxlik <- function(tau, lambda) {
  function(dat) {
    loglik <- function(params) {
      alpha <- params['alpha']
      beta <- params['beta']
      y_star <- dat$y_star
      z <- dat$z
      pi_y <- plogis(alpha + beta * z)
      p_y_star <- pi_y * plogis(tau + lambda) + (1 - pi_y) * plogis(tau)
      logp <- dbinom(y_star, 1, prob = p_y_star, log = TRUE) * dat$count
      return(logp)
    }
    res <- maxLik(loglik, start = c('alpha' = 0, 'beta' = 0), method = "BHHH")
    tibble(est_alpha_MLE = coef(res)['alpha'], est_beta_MLE =
coef(res)['beta'], se_alpha_MLE = stdEr(res)['alpha'], se_beta_MLE =
stdEr(res)['beta'])
  }
}

estimator_misclassification_maxlik <-
get_estimator_misclassification_maxlik(tau = cond$tau, lambda = cond$lambda)
```

```

dat_samp <- simulate_dataset(cond)
res_mle <- estimator_misclassification_maxlik(dat_samp)
res_naive <- estimator_naive(dat_samp)
data.samp <- simulate_dataset(condition)

estimator_relative_risk_mle <- function(dat, est_alpha_MLE, est_beta_MLE) {
  exp_est_alpha_MLE <- exp(est_alpha_MLE)
  exp_est_beta_MLE <- exp(est_beta_MLE)
  prob_unvaccinated <- exp_est_alpha_MLE / (1 + exp_est_alpha_MLE)
  prob_vaccinated <- (exp_est_alpha_MLE * exp_est_beta_MLE) / (1 +
exp_est_alpha_MLE * exp_est_beta_MLE)
  relative_risk <- prob_vaccinated / prob_unvaccinated
  tibble(est_rr_mle = relative_risk)
}

run_condition <- function(cond, nsim = 1) {
  estimator_misclassification_maxlik_new <-
get_estimator_misclassification_maxlik(tau = cond$tau, lambda = cond$lambda)
  map_df(1:nsim, function(isim) {
    dat_samp <- simulate_dataset(cond)
    res_mle <- estimator_misclassification_maxlik(dat_samp)
    res_naive <- estimator_naive(dat_samp)
    res_rr_mle <- estimator_relative_risk_mle(dat_samp, res_mle$est_alpha_MLE,
res_mle$est_beta_MLE)
    bind_cols(res_naive, res_mle, res_rr_mle)
  })
}

dummy_result <- run_condition(conditions[1, ], nsim = 1)
dummy_result[1, ] <- NA

conditions_subset <- conditions

system.time({
  res_overall <- map_df(1:nrow(conditions_subset), function(icond) {
    cat(sprintf("%d\n", icond))
    current_condition <- conditions_subset[icond, ]
    res <- tryCatch(run_condition(current_condition, nsim = nsim), error =
function(e) { dummy_result })
    res %>% bind_cols(current_condition)
  })
})

warnings()
save(res_overall, file = "res_overall.rdata")
head(res_overall)

res_overall <- res_overall %>%
  group_by(n, p_vaccinated, beta_true, alpha_true, p_true_negative,
p_true_positive)

absolute_error <- res_overall %>% summarize(mae_naive =
median(abs(est_beta_glm_naive - beta_true)), mae_MLE = median(abs(est_beta_MLE
- beta_true)), rmse_naive = sqrt(mean( (est_beta_glm_naive - beta_true)^2 )),
rmse_MLE = sqrt(mean((est_beta_MLE - beta_true)^2)))
absolute_error %>%
  pivot_longer(starts_with("mae")) %>%
  mutate(Estimator = gsub("mae_", "", name), `Median absolute error in  $\beta` =
value) %>%
  ggplot(aes(n, `Median absolute error in  $\beta`, color = Estimator)) +
  scale_y_continuous(trans='log10') +
  scale_x_continuous(breaks = c(200, 1000, 2000, 10000)) +
  facet_grid(alpha_true ~ p_true_negative + p_true_positive, scales = "free_y")
+
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +$$ 
```

```

  theme(axis.text.x = element_text(angle = 90, size = 6, vjust = 1, hjust=1),
legend.position="top") +
  geom_point() + geom_line() +
  ggtitle("Median absolute error in estimates of  $\beta$ ")

mse_error <- res_overall %>% summarize(mse_naive = mean((est_beta_glm_naive -
beta_true)^2), mse_MLE = mean((est_beta_MLE - beta_true)^2), rmse_naive =
sqrt(mean( (est_beta_glm_naive - beta_true)^2 )), rmse_MLE =
sqrt(mean((est_beta_MLE - beta_true)^2)))
mse_error %>%
  pivot_longer(starts_with("mse")) %>%
  mutate(Estimator = gsub("mse_", "", name), `Mean squared error in  $\beta$ ` = value)
%>%
  ggplot(aes(n, `Mean squared error in  $\beta$ `, color = Estimator)) +
  scale_y_continuous(trans='log10') +
  scale_x_continuous(breaks = c(200, 1000, 2000, 10000)) +
  facet_grid(alpha_true ~ p_true_negative + p_true_positive, scales = "free_y")
+
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +
  theme(axis.text.x = element_text(angle = 90, size = 6, vjust = 1, hjust=1),
legend.position="top") +
  geom_point() + geom_line() +
  ggtitle("Mean squared error in estimates of  $\beta$ ")

res_overall %>% summarize(across(starts_with("est"), mean, na.rm = TRUE)) %>%
  ungroup %>% select(n, starts_with("est"))

ests_median <- res_overall %>% summarize(across(starts_with("est"), median,
na.rm = TRUE))
head(ests_median)

ests_median %>%
  pivot_longer(c("est_beta_glm_naive", "est_beta_MLE")) %>%
  mutate(Estimator = gsub("est_beta.*_", "", name)) %>%
  mutate(bias = value - beta_true, abs_bias = abs(bias), pct_relbias = 100 *
abs_bias / beta_true) %>%
  ggplot(aes(n, bias, color = Estimator)) +
  scale_x_continuous(breaks = unique(ests_median$n), trans = "sqrt") +
  facet_grid(alpha_true ~ p_true_negative + p_true_positive) +
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +
  theme(axis.text.x = element_text(angle = 90, size = 8), legend.position="top")
+
  geom_hline(yintercept = 0, linetype = 2) +
  geom_point() + geom_line() +
  ggtitle("Bias in estimates of  $\beta$ ") +
  coord_cartesian(ylim = c(NA, 1))

res_sd <- res_overall %>%
  summarize(across(starts_with("est"), mad, na.rm = TRUE)) %>%
  ungroup %>%
  bind_cols(
    res_overall %>%
      summarize(across(starts_with("se"), median, na.rm = TRUE)) %>%
      ungroup %>% select(starts_with("se"))
  )

res_sd %>%
  mutate(`Relative efficiency` = est_beta_MLE/est_beta_glm_naive) %>%
  ggplot(aes(n, `Relative efficiency`, color = factor(alpha_true))) +
  scale_x_continuous(breaks = unique(ests_median$n), trans = "sqrt") +
  scale_y_continuous(trans = "log10") +
  facet_grid(p_true_negative ~ p_true_positive, scales = "free_y") +
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +

```

```

  theme(axis.text.x = element_text(angle = 90, size = 8), legend.position="top")
+
  geom_hline(yintercept = 1, linetype = 2) +
  geom_point() + geom_line() +
  ggtitle("How much more variable is the MLE?")

res_sd %>%
  mutate(`se:sd ratio` = se_beta_MLE / est_beta_MLE) %>%
  ggplot(aes(n, `se:sd ratio`, color = factor(alpha_true))) +
  scale_x_continuous(breaks = unique(ests_median$n), trans = "sqrt") +
  scale_y_continuous(trans = "log10") +
  facet_grid(p_true_negative ~ p_true_positive, scales = "free_y") +
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +
  theme(axis.text.x = element_text(angle = 90, size = 8), legend.position="top")
+
  geom_hline(yintercept = 1, linetype = 2) +
  geom_point() + geom_line() +
  ggtitle("How accurate are MLE-estimated standard errors?")

# Calculate mean squared error
mse_error <- res_overall %>%
  summarise (
    mse_naive = mean((est_rr_naive - beta_true)^2),
    mse_MLE = mean((est_rr_mle - beta_true)^2),
    rmse_naive = sqrt(mean((est_rr_naive - beta_true)^2)),
    rmse_MLE = sqrt(mean((est_rr_mle - beta_true)^2))
  )

# Plot mean squared error
mse_error %>%
  pivot_longer(starts_with("mse")) %>%
  mutate(
    Estimator = gsub("mse_", "", name),
    MeanSquaredErrorInBeta = value
  ) %>%
  ggplot(aes(n, MeanSquaredErrorInBeta, color = Estimator)) +
  scale_y_continuous(trans='log10', limits = c(-1, 5.0)) +
  scale_x_continuous(breaks = c(200, 1000, 2000, 10000)) +
  facet_grid(alpha_true ~ p_true_negative + p_true_positive, scales = "free_y")
+
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +
  theme(axis.text.x = element_text(angle = 90, size = 6, vjust = 1, hjust=1),
legend.position="top") +
  geom_point() + geom_line() +
  ggtitle("Mean squared error in estimates of Relative Risk")

# Mean of estimates over replications per condition
mean_estimates <- res_overall %>%
  summarize(across(starts_with("est"), mean, na.rm = TRUE)) %>%
  ungroup() %>% select(n, starts_with("est"))

# Median of estimates
ests_median <- res_overall %>%
  summarize(across(starts_with("est"), median, na.rm = TRUE))
head(ests_median)

# Plot bias in estimates of relative risk
ests_median %>%
  pivot_longer(c("est_rr_naive", "est_rr_mle")) %>%
  mutate(
    Estimator = gsub("est_rr.*_", "", name),
    bias = value - beta_true,
    abs_bias = abs(bias),
    pct_relbias = 100 * abs_bias / beta_true
  )

```

```
) %>%
  ggplot(aes(n, bias, color = Estimator)) +
  scale_x_continuous(breaks = unique(ests_median$n), trans = "sqrt") +
  facet_grid(alpha_true ~ p_true_negative + p_true_positive) +
  ggthemes::theme_few() +
  ggthemes::scale_color_colorblind() +
  theme(axis.text.x = element_text(angle = 90, size = 8), legend.position="top")
+
  geom_hline(yintercept = 0, linetype = 2) +
  geom_point() + geom_line() +
  ggtitle("Bias in Estimates of Relative Risk") +
  coord_cartesian(ylim = c(NA, 4))
```

Appendix 5: Bias Analysis for the Estimated Relative Risk

The patterns of bias observed for the relative risk closely mirrored those seen for the log odds ratio, with the intensity of fluctuations growing together with the level of specificity. These fluctuations became more distinct at lower prevalence rates. In these simulations, the naïve produced consistent values.

Under high prevalence conditions, when the true alpha was -0.5 and specificity was 0.99, the naïve estimator and MLE demonstrated certain degrees of bias in estimating relative risk, especially when the quantified effect size was relatively low such as 0.1 (**Fig 4**). However, when we raised the true beta value to 0.5, the MLE exhibited less bias than the naïve estimator (**Fig 5**). This was evidenced by point estimates nearing zero in simulations with a sample size of 10,000, suggesting that the MLE might show less bias within the relative risk than with the log of odds ratio.

When the strength of association was raised to the highest value of 0.9, the naïve estimator remained steady across simulations, producing consistently stable estimates for bias in most plots (**Fig 6**). On the other hand, in comparison, the MLE showed an apparent decrease in bias, especially in scenarios with high prevalence and high specificity (0.99). This was particularly noticeable for the highest beta coefficient (0.9), suggesting that integrating misclassification into our model can be beneficial. This suggests that estimate bias could potentially be reduced in scenarios in which misclassification could be appropriately adjusted for.

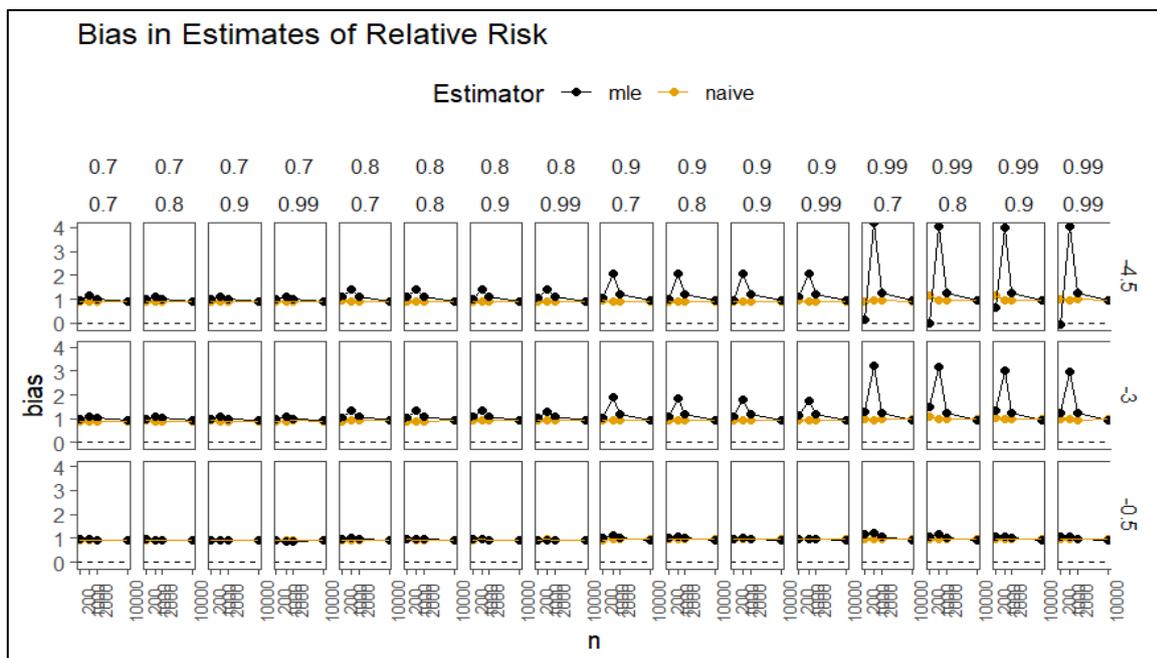


Figure 4: Assessment of Bias of the Relative Risk under the Assumption of True Beta Value Equalling 0.1

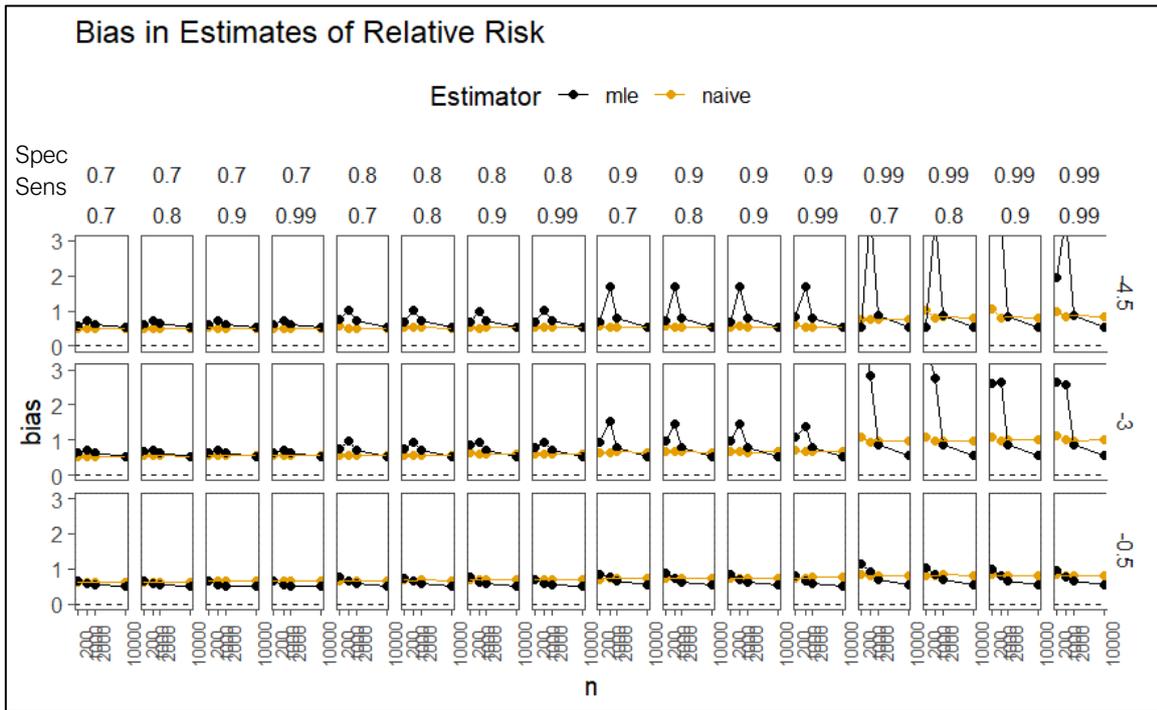


Figure 5: Assessment of Bias of the Relative Risk under the Assumption of True Beta Value Equalling 0.5.

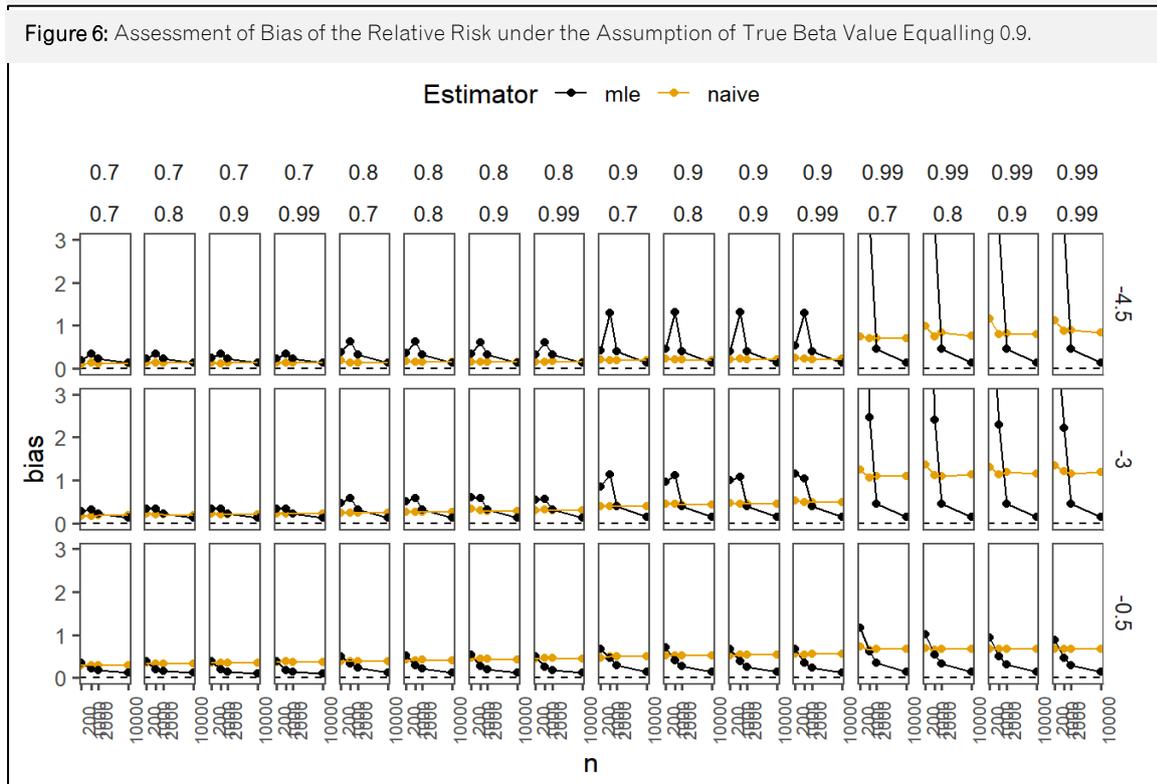


Figure 6: Assessment of Bias of the Relative Risk under the Assumption of True Beta Value Equalling 0.9.

Appendix 6: Mean Squared Error Analysis for the Estimated Relative Risk

When true beta values were set to 0.1, we observed a systematic increase in the slope's steepness across simulation runs for the MSE (**Fig 10**). As expected, these estimates became less dramatic in scenarios of high prevalence and low specificity and sensitivity. The most apparent effect is observed when alpha true is set to -4.5 and specificity and sensitivity parameters are approximately 0.99. Interestingly, in this context, there is a negligible distinction between the MSE estimates derived from the naïve and MLE methods.

When we adjusted the beta true to 0.5, an evident pattern emerged (**Fig 11**), the MLE consistently produced lower error estimates across nearly all simulated scenarios. This trend became even more pronounced when we further increased beta true to 0.9 (**Fig 12**). From these observations, it could be inferred that the MLE is particularly effective in minimising error for the relative risk when the underlying relationship is strong. This is in line with theoretical expectations and provides valuable insights into the comparative performance of naïve and MLE estimation methods under varying conditions.

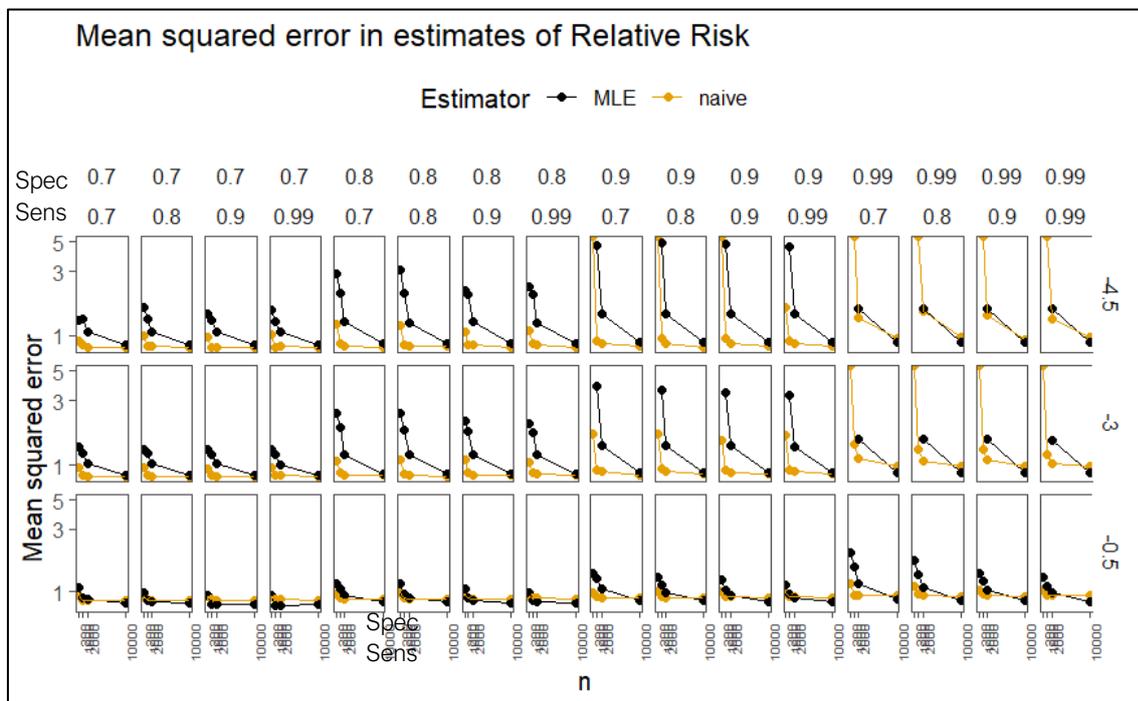


Figure 10: Assessment of MSE of the Relative Risk under the Assumption of True Beta Value Equalling 0.1.

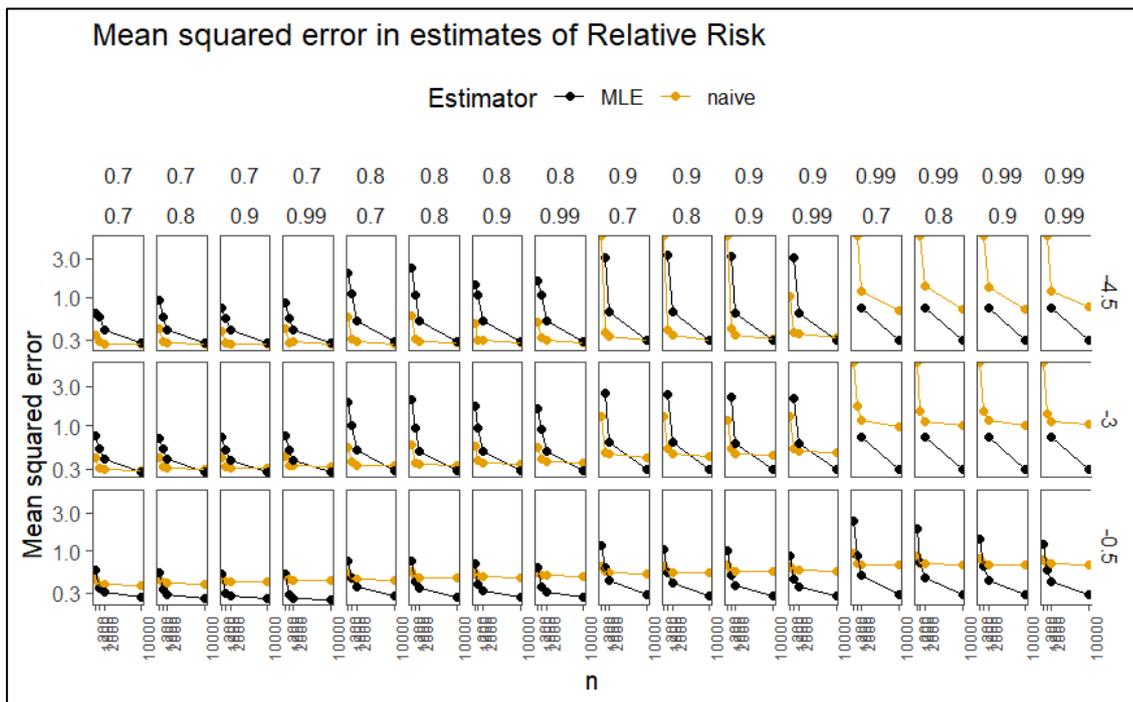


Figure 11: Assessment of MSE of the Relative Risk under the Assumption of True Beta Value Equalling 0.5.

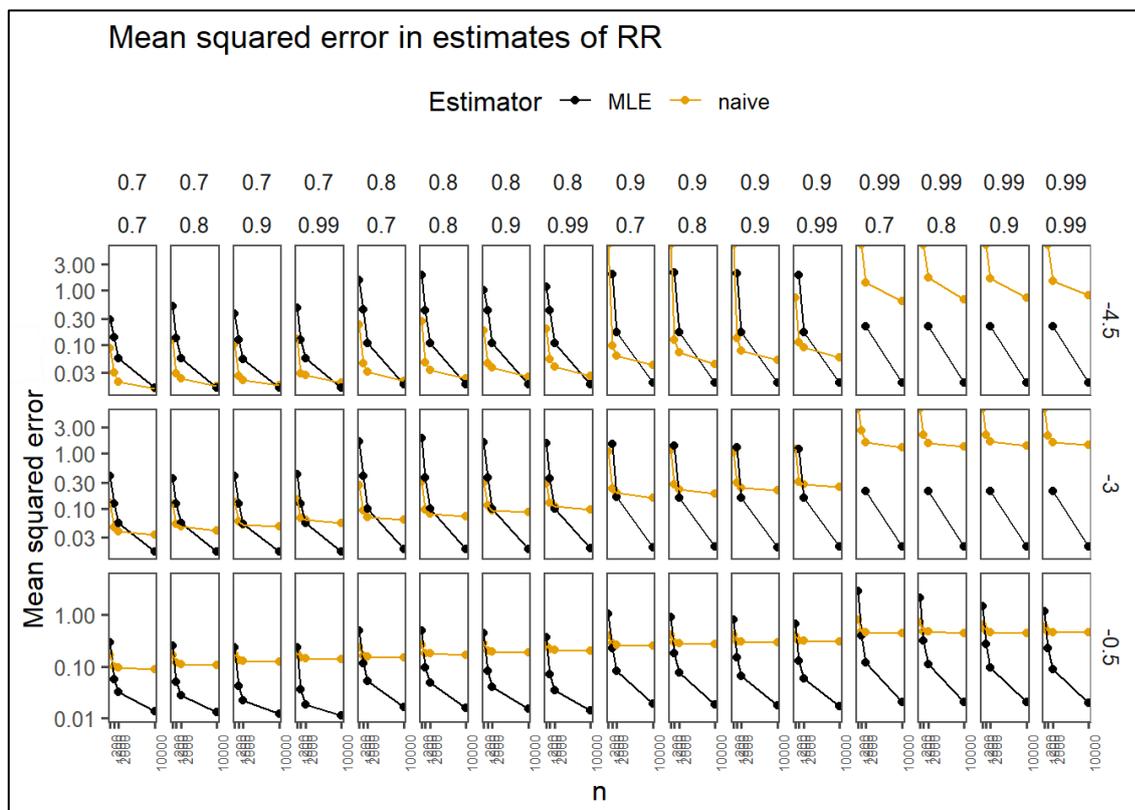


Figure 12: Assessment of MSE of the Relative Risk under the Assumption of True Beta Value Equalling 0.9.